PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/53, 9/02, C12Q 1/32, C07K 16/40 — A1

(11) International Publication Number: WO 95/28489

(43) International Publication Date: 26 October 1995 (26.10.95)

(21) International Application Number: PCT/US95/04567

(22) International Filing Date: 13 April 1995 (13.04.95)

(30) Priority Data:
08/227,357 13 April 1994 (13.04.94) US

(71) Applicant: THE UAB RESEARCH FOUNDATION [US/US]; Suite 1120G, 701 South 20th Street, Birmingham, AL 35294-0111 (US).

(72) Inventors: DIASIO, Robert, B.; 1225 Branchwater Lane, Birmingham, AL 35216 (US). LU, Zhihong; 1824 Russet Woods Lane, Birmingham, AL 35244 (US). ZHANG, Ruiwen; 1824 Russet Woods Lane, Birmingham, AL 35244 (US). JOHNSON, Martin; 211 Coral Circle, Alabaster, AL 35007 (US). CHENG, Xiaogang; 840 - 16th Street South, Birmingham, AL 35205 (US).

(74) Agent: WILSON, Mark, B.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210-4433 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.

(54) Title: DIHYDROPYRIMIDINE DEHYDROGENASE COMPOSITIONS AND METHODS OF USE

(57) Abstract

Disclosed are methods and compositions for use in detecting and quantifying the enzyme dihydropyrimidine dehydrogenase (DPD) for use in, e.g., optimizing 5-fluorouracil doses given to cancer patients. Particularly described are antibodies, including monoclonal antibodies, to the human form of DPD; DNA sequences from bovine and human DPD; immunological and molecular biological means by which to detect DPD; and methods of designing effective cancer treatment strategies based upon information gained concerning DPD levels. Also disclosed is molecular characterization of a genetic lesion leading to DPD deficiency in humans and diagnostic methods for genetic screening of this mutation for patients undergoing FUra treatment.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

DESCRIPTION

DIHYDROPYRIMIDINE DEHYDROGENASE COMPOSITIONS AND METHODS OF USE

A. Field of the Invention

The present invention relates generally to DNA sequences encoding mammalian dihydropyrimidine dehydrogenase (DPD); immunological and molecular biological methods of detecting DPD; and detection and quantitation of the DPD enzyme, as may be employed for optimizing 5-fluorouracil (FUra) doses, and determining increased FUra toxicity in DPD-deficient subjects. More particularly, it concerns the first molecular cloning of mammalian DPD; antibodies, including monoclonal antibodies from hybridomas, to the human form of DPD; methods of making and using such antibodies; methods of designing effective cancer treatment strategies based upon information gained concerning DPD levels; and diagnostic screening methods for determining genetic susceptibility to DPD deficiency.

B. Description of the Related Art

1. Dihydropyrimidine Dehydrogenase (DPD)

Dihydropyrimidine dehydrogenase (EC 1.3.1.2, dihydrouracil dehydrogenase, dihydrothymine dehydrogenase, DPD) catalyzes the initial and rate-limiting step in pyrimidine catabolism; the reduction of pyrimidines to 5,6-dihydropyrimidines (Traut and Loechel, 1984). Dihydrouracil and dihydrothymine are further catabolized to β-alanine or β-aminoisobutyric acid respectively with release of CO_2. DPD is not only important in regulating systemic levels of uracil and thymine (Wasternak, 1980), but also in the synthesis of β-alanine (Matthews *et al.*, 1992).

DPD has been partially purified and characterized from liver of several mammals including human (Lu *et al.*, 1992), bovine (Lu *et al.*, 1993), rat (Shiotani and Weber, 1981, and pig (Podschun *et al.*, 1989). These studies have demonstrated that DPD is a complex enzyme consisting of two identical subunits, containing FMN, FAD, and iron-sulfur centers, and utilizing NADPH as a cofactor (Lu *et al.*, 1992; Lu *et al.*, 1993; Shiotani and Weber, 1981; Podschun *et al.*, 1989). Availability of the purified enzyme has permitted preparation of polyclonal antibody (Lu *et al.*, 1992; Podschun *et al.*, 1989) and determination of the amino acid sequence (Lu *et al.*, 1992; Porter *et al.*, 1992a; 1992b),

which in turn has enabled cloning of the cDNA for DPD (Johnson *et al.*, 1995; Yokota *et al.*, 1994).

Since the major site of pyrimidine catabolism is in the liver (Ho *et al.*, 1986), most of the studies involving DPD have been performed using liver tissue. In the last three decades, DPD has been purified to varying degrees from liver of several species, including cow (Grisolia *et al.*, 1957; Porter *et al.*, 1992a), rat (Fritzson, 1960; Shiotani and Weber, 1981; Fujimoto *et al.*, 1990), mouse (Sanno *et al.*, 1970), and pig (Goedde *et al.*, 1970; Podschun *et al.*, 1989; Podschun *et al.*, 1990). A major limitation of this work, however, was the fact that homogeneity in purification was not obtained for most of these DPD preparations.

Prior to the work of the present inventors, very little was known about the human liver enzyme. Studies have suggested that DPD might be species-specific since the antiserum to rat liver DPD does not precipitate dog or guinea pig liver DPD activity (Fujimoto et al., 1990). Species differences in this enzyme were also shown in a recent report of pig liver DPD (Podschun *et al.*, 1989) compared with rat liver DPD (Shiotani and Weber, 1981; Fujimoto, *et al.*, 1990). In the past, limited successes with purification of the human enzyme have prevented the detailed analysis of DPD. Moreover, the lack of DNA sequences encoding human DPD made it impossible to study the expression or regulation of DPD *in vivo*.

2. DPD and 5-Fluorouracil (FUra) Catabolism

It was shown that in addition to its normal biological activity, DPD also catalyzes the reduction of various pyrimidine analogs, including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra). Studies have demonstrated that more than 85% of administered FUra, one of the most frequently used anticancer drugs, is catabolized by DPD (Diasio and Harris, 1989). It has also been demonstrated that the anticancer efficacy of FUra is related to DPD activity (Iigo *et al.*, 1989). Because of its role in FUra catabolism, DPD activity is, therefore, critical in regulating the availability of FUra once administered to a patient, and hence, governs FUra's effect as an anticancer therapeutic. Moreover, DPD is critical in the balance between effective FUra levels *in vivo* for therapy and limiting FUra toxicity once administered. Recently there have been several reports of a

pharmacogenetic disorder in which individuals with decreased DPD activity developed life-threatening toxicity following exposure to FUra.

Experimental and clinical studies have demonstrated DPD activity to have a circadian variation (Harris *et al.*, 1988, 1990; Daher *et al.*, 1991). This circadian pattern may have an important role in FUra chemotherapy, since FUra plasma levels have a corresponding inverse circadian pattern in patients receiving FUra chemotherapy (Harris *et al.*, 1990). Additional studies with competitive DPD inhibitors (Wasternack, 1980; Martin *et al.*, 1978; Daher *et al.*, 1991) have also shown the importance of this enzyme in cancer chemotherapy.

3. FUra Toxicity in DPD-Deficient Patients

Individuals with absent or significantly decreased lymphocytic DPD activity may develop life-threatening toxicity following exposure to FUra (Diasio *et al.*, 1988; Lu *et al.*, 1993). Since the initial reports several years ago, there have been an increasing number of cases described suggesting that this disorder may be more frequent than initially thought (Lu *et al.*, 1993; Diasio and Lu, 1994). Deficient individuals have been identified using both direct measurement of DPD activity (Diasio *et al.*, 1988; Lu *et al.*, 1993; Diasio and Lu, 1994; Harris *et al.*, 1991) as well as quantitation of DPD protein by western blot analysis (Diasio *et al.*, 1994; Zhang *et al.*, 1994).

The observation of an inherited (pharmacogenetic) disorder in which individuals with absent or significantly decreased DPD activity develop life-threatening toxicity following exposure to FUra has heightened interest in the biochemical and molecular basis for altered enzyme activity (Diasio *et al.*, 1988; Lu *et al.*, 1993).

4. DNA Segments Encoding DPD

Despite the significant interest in DPD, relatively little has been previously known about the gene which encodes DPD, its sequence, or its regulation or factors which govern its expression. Likewise no genetic screening methods currently exist for determination of genetic defects resulting in DPD enzyme deficiency. Prior to the work of the present inventors no genetic lesion(s) had been identified which contributed to decreased DPD activity in vivo and concomitant increased FUra sensitivity.

5. Deficiencies in the Prior Art

Because it is unlikely that DPD obtained from non-human tissues could be used to investigate the biology and molecular genetics of the human enzyme and the gene(s) which encode it, it is highly desirable to purify the human enzyme and characterize it in more detail. Likewise, the development of effective genetic screening methods and studies of the genetic regulation of DPD activity *in vivo* are impossible without elucidation of the DNA sequence of nucleic acids which encode human DPD.

Moreover, there is currently no effective means by which to detect and quantitate DPD levels in patient samples. All the available methods for analyzing DPD, such as complicated enzyme assays, are slow and labor-intensive, and not generally adaptable for patient screening. The development of methods by which patient DPD deficiencies could be readily detected would be a significant advance that would allow adjustment of the dose of FUra prior to administration of the drug.

In addition to detecting individuals who are deficient in DPD activity due to genetic factors, DPD activity may also be altered by other factors. Thus, monitoring DPD activity before using fluoropyrimidine drugs in general is considered to be very important. In light of the above, it is clear that there remains in the art a need for improved assays to quantitate DPD which would allow better therapeutic use of the widely used, but partially toxic, drug FUra. Finally, the availability of genetic screening methods and identification of genetic lesion(s) in DPD-deficient patients would greatly facilitate the treatment of proliferative cell disorders using FUra and related chemotherapeutics, and forward the areas of medical arts particularly in the area of anticancer therapies.

SUMMARY OF THE INVENTION

The present invention seeks to overcome these and other drawbacks inherent in the prior art by providing methods and compositions for use in detecting and quantifying the human DPD enzyme and, therefore, for optimizing the FUra doses given to certain cancer patients. In particular, the invention provides DNA sequence encoding mammalian DPD. Moreover, the invention concerns the isolation and characterization of nucleic acid segments encoding bovine liver and human lymphocyte DPD, determination of the amino

acid sequences of bovine and human DPD, and identification of a frameshift mutation in the DNA encoding DPD in a DPD-deficient patient sensitive to FUra, and methods and compositions for assaying patients for increased FUra sensitivity.

The present invention also describes a novel purification procedure for human liver DPD, and includes new information on the properties of mammalian DPD. Furthermore, the invention makes available for the first time pure, human DPD enzyme, a polyclonal antibody against this enzyme, and new data on amino acid composition and sequence to provide a molecular basis for further biochemical and molecular analyses of this enzyme, particularly relevant to human DPD activity, DPD deficiency, and molecular methods for treating cancer patients with drugs such as FUra.

Because attempts to purify DPD from human liver using methods of the prior art were unsuccessful in purifying the mammalian enzyme to homogeneity (Shiotani and Weber, 1981; Podschun *et al.*, 1990), novel methods are disclosed with provide a purified human DPD. The present invention overcomes this limitation by providing isolation and characterization of the DPD from bovine liver and human lymphocytes.

An aspect of the invention concerns the complete amino acid sequencing of the bovine liver DPD (disclosed in SEQ ID NO:2), and the complete amino acid sequencing of human DPD (disclosed in SEQ ID NO:4). Additional aspects of the invention identify the complete nucleic acid sequences encoding both bovine (SEQ ID NO:1) and human (SEQ ID NO:3) DPD. Human lymphocytes are typically used for assessing DPD activity (Diasio *et al.*, 1988; Lu *et al.*, 1993; Diasio and Lu, 1994; Harris *et al.*, 1991). Because previous studies of lymphocyte and liver DPD have suggested the possibility of more than one isozyme (Naguib *et al.*, 1985), the cDNAs for both bovine liver DPD and human lymphocytes were obtained.

A surprising aspect of the invention concerns the cDNA sequence of DPD from an individual with increased FUra sensitivity. Unlike the normal DPD cDNA obtained from individuals having normal DPD activity, the individual showing an increased sensitivity to FUra had an altered DPD gene.

In certain aspects, the invention concerns methods for determining a therapeutically effective dose of FUra for administration to a patient, such as a cancer

patient, comprising determining the amount of dihydropyrimidine dehydrogenase (DPD) present within a biological sample from said patient and adjusting the dose of FUra to be administered according to the amount of DPD detected.

Because the present invention provides the complete nucleotide sequences of bovine liver and human lymphocyte DPDs (together with their completely translated amino acid sequences), elucidation of tertiary structures of these enzymes is now possible. Moreover, cofactor binding, specific interactions with inactivators of DPD, and molecular analyses involving this critical enzyme is now made available.

The present invention also provides complete cDNAs for DPD which in turn provides insight into the molecular basis of the altered DPD activity observed with the inherited pharmacogenetic disorder of increased FUra toxicity and DPD deficiency.

More particularly, where an increased amount of DPD is detected in a patient's sample, one would generally increase the dose of FUra administered to the patient. Equally, where an increased amount of DPD is detected in a patient's sample, one may administer to said patient FUra is combination with an agent to inhibit DPD, such as, *e.g.*, uridine, 5-ethynyluracil (EU), interferon, leucovorin, cimetidine (CMT) or 5-benzyloxy-benzyluracil (BBU).

Where a decreased amount of DPD is detected in a patient's sample, one would generally decrease the dose of FUra administered to the patient. Upon detecting a significantly decreased amount of DPD, one would either not administer any FUra to the patient, or one may administer FUra is combination with the DPD enzyme itself, such as purified human, bovine or rat DPD.

The present invention provides immunoassays for detecting the amount of DPD in biological samples, such as body fluids (*e.g.*, blood and plasma). The immunoassays may employ polyclonal antibodies to DPD, but most preferably, will employ a monoclonal antibody that has binding affinity for human DPD.

The invention also contemplates the use of molecular biological methods to detect DPD, wherein the patient's DPD levels are determined by means of determining the amount of a nucleic acid that encodes DPD present within a biological sample from the patient. To conduct such a method, one would contact nucleic acids from the biological

sample with a DNA segment that encodes mammalian DPD, under conditions effective to allow hybridization of substantially complementary nucleic acid sequences, and then detect the complex of complementary nucleic acids thus formed.

Also provided are methods for determining an appropriate, or a more or a most appropriate, mode of treatment for a cancer patient, which methods comprise identifying a patient with a deficiency in the amount of DPD and treating the patient by a method other than only administering FUra alone to the patient. Examples include both the complete avoidance of FUra treatment and administering FUra in combination with DPD. The detection means may be both immunoassays using, *e.g.,* monoclonal antibodies that have binding affinity for human DPD; and molecular biological assay using a nucleic acid segment or segments, either DNA or RNA, that encode mammalian DPD.

To detect DPD using the invention one may contact a sample suspected of containing DPD with a first monoclonal antibody that binds to human DPD, under conditions effective to allow the formation of immune complexes, and then detecting the immune complexes so formed. This method may employ a first antibody that is linked to a detectable label thus allowing the immune complexes to be detected by detecting the presence of the label. Also, the immune complexes may be detected by means of a second antibody linked to a detectable label, the second antibody having binding affinity for the first antibody.

The invention also provides monoclonal antibodies that have binding affinity for human DPD. Such monoclonal antibodies may be those that are obtainable by immunizing an animal with DPD purified from human liver, in an amount effective to stimulate the generation of B cells producing antibodies specific for DPD, and then immortalizing such B cells and obtaining a monoclonal antibody secreted by the immortalized B cells. Hybridomas that produce such monoclonal antibodies are also encompassed by the invention.

One may prepare a suitable DPD antigen preparation, such as a purified human liver DPD preparation, by following the methods described herein. Monoclonal antibody generation may be achieved by using methods that will be well known to those of skill in

the art in light of the present disclosure, *e.g.*, as outlined in the description of the preferred embodiments.

Immunodetection kits that comprise, in a suitable container, a first monoclonal antibody that binds to human DPD and an immunodetection reagent are another aspect of the invention. Such kits may use an immunodetection reagent that is a detectable label linked to the first antibody itself, or an immunodetection reagent that is a detectable label that is linked to a second antibody that has binding affinity for the first antibody.

Still further aspects of the invention are the identification of a specific frameshift mutation in the DNA segment encoding mammalian DPD. This frameshift mutation results in DPD deficiency in cells containing and expressing this DNA segment. Methods are disclosed which provide the first diagnostic screening for genetic DPD deficiency in an animal. The methods employ the identification of the frameshift mutation disclosed herein. The DNA segments that include nucleic acid sequences as set forth in either SEQ ID NO:1 or SEQ ID NO:3 are particularly preferred in methods relating to genetic detection of DPD deficiency, and for use as an anti-proliferative agents such as in the treatment of cancer.

1. Generating an Immune Response to DPD

The present invention thus also provides methods of generating an immune response, which methods generally comprise administering to an animal, including a human subject, a pharmaceutically acceptable composition comprising an immunologically effective amount of a DPD protein or peptide composition. The composition may include partially or significantly purified DPD proteins or peptides, obtained from natural or recombinant sources, which proteins or peptides may be obtainable from human, bovine, or recombinant bacterial sources. Smaller peptides that include reactive epitopes, such as those between about 30 and about 50 amino acids in length will often be preferred.

By "immunologically effective amount" is meant an amount of a DPD protein or peptide composition that is capable of generating an immune response in the recipient animal. This includes both the generation of an antibody response (B-cell response), and/or the stimulation of a cytotoxic immune response (T-cell response). The generation of such an immune response will have utility in both the production of useful bioreagents,

e.g., cytotoxic T-lymphocytes (CTLs) and, more particularly, reactive antibodies, for use in diagnostic embodiments, and will also have utility in various prophylactic or therapeutic embodiments. Therefore, although these methods for the stimulation of an immune response include vaccination regimens designed to prevent or lessen significant FUra toxicity and/or DPD deficiencies, and treatment regimens that may lessen the severity or duration of any DPD deficiency or FUra toxicity, it will be understood that achieving either of these end results is not necessary for practicing these aspects of the invention.

Another means contemplated by the inventors for generating an immune response in an animal includes administering to the animal, or human subject, a pharmaceutically acceptable composition comprising an immunologically effective amount of a DPD-encoding nucleic acid composition (*i.e.*, an amount capable of stimulating a B cell and/or T cell response). The stimulation of specific antibodies and CTL responses upon administering to an animal a nucleic molecule is now well known in the art (Tang *et al.*, 1992; Cox *et al.*, 1993; Fynan *et al.*, 1993; Ulmer *et al.*, 1993; Wang *et al.*, 1993; and Whitton *et al.*, 1993).

This technology, often referred to as genetic immunization, is contemplated to be particularly suitable to protect against viral infections. Indeed, immunization with DNA has been successfully employed to protect animals from challenge with influenza A (Ulmer *et al.*, 1993). Therefore, the use of the DPD-encoding nucleic acid compositions of the present invention in techniques such as those described (Ulmer *et al.*, 1993; incorporated herein by reference), is considered to be particularly useful as a vaccination regimen. The DPD-encoding DNA segments could be used in virtually any form, including naked DNA and plasmid DNA, and may be administered to the animal in a variety of ways, including parenteral, mucosal and gene-gun inoculations, as described (see for example, Fynan *et al.*, 1993).

Immunoformulations of this invention, whether intended for vaccination, treatment, or for the generation of antibodies useful in DPD detection, may comprise whole DPD proteins or antigenic peptide fragments from these proteins. As such, antigenic functional equivalents of the proteins and peptides described herein also fall within the scope of the present invention. An "antigenically functional equivalent" protein or peptide is one that

incorporates an epitope that is immunologically cross-reactive with one or more epitopes of the DPD proteins. Antigenically functional equivalents, or epitopic sequences, may be first designed or predicted and then tested, or may simply be directly tested for cross-reactivity.

Suitable competition assays that may be employed include protocols based upon immunohistochemical assays, ELISAs, RIAs, Western or dot blotting and the like. In any of the competitive assays, one of the binding components, generally the known element, such as the DPD protein or peptide, or the known antibody, such as the monoclonal antibody, will be labeled with a detectable label and the test components, that generally remain unlabeled, will be tested for their ability to reduce the amount of label that is bound to the corresponding reactive antibody or antigen.

As an exemplary embodiment, to conduct a competition study between DPD and any test antigen, one would first label DPD with a detectable label, such as, *e.g.*, biotin or an enzymatic, radioactive or fluorogenic label, to enable subsequent identification. One would then incubate the labelled antigen with the other, test, antigen to be examined at various ratios (*e.g.*, 1:1, 1:10 and 1:100) and, after mixing, one would then add the mixture to a known antibody, such as anti-DPD. Preferably, the known antibody would be immobilized, *e.g.*, by attaching to an ELISA plate. The ability of the mixture to bind to the antibody would be determined by detecting the presence of the specifically bound label. This value would then be compared to a control value in which no potentially competing (test) antigen was included in the incubation.

The assay may be any one of a range of immunological assays based upon hybridization, and the reactive antigens would be detected by means of detecting their label, *e.g.*, using streptavidin in the case of biotinylated antigens or by using a chromogenic substrate in connection with an enzymatic label or by simply detecting a radioactive or fluorescent label. An antigen that binds to the same antibody as anti-DPD, for example, will be able to effectively compete for binding to DPD and thus will significantly reduce DPD binding, as evidenced by a reduction in the amount of label detected.

The reactivity of the labeled antigen, *e.g.*, DPD, in the absence of any test antigen would be the control high value. The control low value would be obtained by incubating the labeled antigen with an excess of unlabeled DPD antigen, when competition would occur and reduce binding. A significant reduction in labeled antigen reactivity in the presence of a test antigen is indicative of a test antigen that is "cross-reactive", *i.e.*, that has binding affinity for the same antibody.

Particular techniques for preparing antibodies in accordance with the invention are disclosed herein. However, it is proposed by the inventors that any of the current techniques known in the art for the preparation of antibodies in general may be employed, through the application of either monoclonal or polyclonal technology, and as represented by the generation of the monoclonal antibody against DPD. Antibodies that are cross-reactive with DPD are also encompassed by the invention, as may be identified by employing a competition binding assay, such as those described above in terms of antigen competition.

Antibodies of the invention may also be linked to a detectable label, such as a radioactive, fluorogenic or a nuclear magnetic spin resonance label. Biolabels such as biotin and enzymes that are capable of generating a colored product upon contact with a chromogenic substrate are also contemplated. Exemplary enzyme labels include alkaline phosphatase, hydrogen peroxidase and glucose oxidase enzymes.

In still further embodiments, the present invention concerns immunodetection methods and associated kits. It is contemplated that the DPD proteins or peptides of the invention may be employed to detect antibodies having reactivity therewith, or, alternatively, antibodies prepared in accordance with the present invention, *e.g.*, anti-DPD or an antibody against DPD-like peptides, may be employed to detect DPD proteins or peptides. Either type of kit may be used in the immunodetection of DPD, present within clinical samples, such as for example in determining the activity of DPD in a patient that is undergoing FUra treatment for proliferative cell disorders. The kits may also be used in antigen or antibody purification, as appropriate.

In general, immunodetection methods will include first obtaining a sample suspected of containing such a protein, peptide or antibody, such as a biological sample

from a patient, and contacting the sample with a first antibody that binds to a DPD protein or peptide, as the case may be, under conditions effective to allow the formation of an immunocomplex (primary immune complex). One then detects the presence of any primary immunocomplexes that are formed.

Contacting the chosen sample with the DPD antibody, under conditions effective to allow the formation of (primary) immune complexes is generally a matter of simply adding the antibody composition to the sample. One then incubates the mixture for a period of time sufficient to allow the added antibodies to form immune complexes with, *i.e.*, to bind to, antigens present within the sample. After this time, the sample composition, such as a tissue section, ELISA plate, dot blot or western blot, will generally be washed to remove any non-specifically bound antigen or antibody species, allowing only those specifically bound species within the immune complexes to be detected.

The detection of immunocomplex formation is well known in the art and may be achieved through the application of numerous approaches known to the skilled artisan and described in various publications, such as, *e.g.*, Nakamura *et al.* (1987), incorporated herein by reference. Detection of primary immune complexes is generally based upon the detection of a label or marker, such as a radioactive, fluorigenic, biological or enzymatic label, with enzyme tags such as alkaline phosphatase, horseradish peroxidase and glucose oxidase being suitable. The antigen (*e.g.*, DPD) or DPD antibody employed may itself be linked to a detectable label, wherein one would then simply detect this label, thereby allowing the amount of bound antigen or antibody present in the composition to be determined.

In one alternative, the primary immune complexes may be detected by means of a second binding ligand that is linked to a detectable label and that has binding affinity for the first protein, peptide or antibody. The second binding ligand is itself often an antibody, which may thus be termed a "secondary" antibody. The primary immune complexes are contacted with the labeled, secondary binding ligand, or antibody, under conditions effective and for a period of time sufficient to allow the formation of secondary immune complexes. The secondary immune complexes are then generally

washed to remove any non-specifically bound labeled secondary antibodies or ligands, and the remaining bound label is then detected.

In yet another alternative, the secondary immune complexes may be detected by means of a tertiary binding ligand that is linked to a detectable label and that has binding affinity for the second binding ligand or antibody. The tertiary binding ligand will again often be an antibody, which may thus be termed a "tertiary" antibody. The secondary immune complexes are contacted with the labeled, tertiary binding ligand, or antibody, under conditions effective and for a period of time sufficient to allow the formation of tertiary immune complexes. The tertiary immune complexes are then generally washed to remove any non-specifically bound labelled antibodies or ligands, and the remaining bound label is then detected.

This latter alternative is exemplified by the currently preferred sandwich ELISA. Here, the DPD antibodies are first immobilized and then contacted with a DPD sample, the secondary antibody is an unlabeled anti-DPD monoclonal antibody, and the tertiary antibody is a commercially available labeled antibody that is specific for a non-variant portion of the second monoclonal antibody.

For diagnostic purposes, it is proposed that virtually any sample suspected of containing either the DPD proteins, peptides or antibodies sought to be detected, as the case may be, may be employed. Exemplary samples include clinical samples obtained from a patient such as blood or serum samples, bronchoalveolar fluid, ear swabs, sputum samples, middle ear fluid or even perhaps urine samples may be employed. Furthermore, it is contemplated that such embodiments may have application to non-clinical samples, such as in the titering of antigen or antibody samples, in the selection of hybridomas, and the like.

In related embodiments, the present invention contemplates the preparation of kits that may be employed to detect the presence of DPD proteins, peptides and/or antibodies in a sample. Generally speaking, kits in accordance with the present invention will include a suitable DPD protein or peptide, or a first antibody that binds to a DPD protein or peptide, together with an immunodetection reagent, and a container for the protein, peptide or antibody and reagent.

The immunodetection reagent will typically comprise a label associated with the protein, peptide or antibody, or associated with a secondary binding ligand. Exemplary ligands might include a secondary antibody directed against the first protein, peptide or antibody, or a biotin or avidin (or streptavidin) ligand having an associated label. Detectable labels linked to antibodies that have binding affinity for a human antibody are also contemplated, *e.g.,* for protocols where the first reagent is a protein that is used to bind to a reactive antibody from a human sample. Of course, as noted above, a number of exemplary labels are known in the art and all such labels may be employed in connection with the present invention. The kits may contain antigen or antibody-label conjugates either in fully conjugated form, in the form of intermediates, or as separate moieties to be conjugated by the user of the kit.

The container will generally include at least one vial, test tube, flask, bottle, syringe or other container, into which the antigen or antibody may be placed, and preferably suitably allocated. Where a second binding ligand is provided, the kit will also generally contain a second vial or other container into which this ligand or antibody may be placed. The kits of the present invention will also typically include a container for the vials in close confinement for commercial sale, such as, *e.g.,* injection or blow-molded plastic containers into which the desired vials are retained.

2. DNA Segments

Important aspects of the present invention concern isolated DNA segments and recombinant vectors encoding DPD, and the creation and use of recombinant host cells through the application of DNA technology, that express DPD. The present invention concerns DNA segments, isolatable from bovine or human, that are free from total genomic DNA and are capable of conferring DPD activity to a recombinant host cell when incorporated into the recombinant host cell. DNA segments capable of conferring DPD activity may encode DPD proteins, peptides, functional domains, *etc.,* and may also be combined with other peptides, cofactors, regulatory proteins, *etc.*

As used herein, the term "DNA segment" refers to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding DPD refers to a DNA segment that contains DPD coding sequences yet is

isolated away from, or purified free from, total genomic DNA of either bovine or human cells. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified DPD gene refers to a DNA segment including DPD coding sequences and, in certain aspects, regulatory sequences, isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, cDNA sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides or peptides.

"Isolated substantially away from other coding sequences" means that the gene of interest, in this case one encoding DPD, forms the significant part of the coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or cDNA coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode DPD that includes within its amino acid sequence the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:4, corresponding to the bovine and human DPD, respectively). Moreover, in other particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode DPD that includes within its amino acid sequence the amino acid sequence of DPD corresponding to human DPD.

In certain embodiments, the invention concerns isolated DNA segments and recombinant vectors that encode a protein or peptide that includes within its amino acid sequence an amino acid sequence essentially as set forth in SEQ ID NO:2 or SEQ ID NO:4. Naturally, where the DNA segment or vector encodes a full length DPD protein, or

is intended for use in expressing the DPD protein, the most preferred sequences are those that are set forth in SEQ ID NO:1 or SEQ ID NO:3 and that encode a protein that retains DPD activity, *e.g.,* as may be determined by the DPD assay, as disclosed herein.

The term "a sequence essentially as set forth in SEQ ID NO:2 or SEQ ID NO:4" means that the sequence substantially corresponds to a portion of SEQ ID NO:2 or SEQ ID NO:4 and has relatively few amino acids that are not identical to, or a biologically functional equivalent of, the amino acids of SEQ ID NO:2 or SEQ ID NO:4. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein. Accordingly, sequences that have between about 70% and about 100%; or more preferably, between about 81% and about 100%; or even more preferably, between about 91% and about 100%; of amino acids that are identical or functionally equivalent to the amino acids of SEQ ID NO:2 or SEQ ID NO:4 will be sequences that are "essentially as set forth in SEQ ID NO:2 or SEQ ID NO:4".

In certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:1 or SEQ ID NO:3. The term "essentially as set forth in SEQ ID NO:1 or SEQ ID NO:3" is used in the same sense as described above and means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:1 or SEQ ID NO:3 and has relatively few codons that are not identical, or functionally equivalent, to the codons of SEQ ID NO:1 or SEQ ID NO:3. Again, DNA segments that encode proteins exhibiting DPD activity will be most preferred. The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine or serine, and also refers to codons that encode biologically equivalent amino acids.

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include

various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, *i.e.,* introns, which are known to occur within genes.

Excepting intronic or flanking regions, and allowing for the degeneracy of the genetic code, sequences that have between about 50% and about 100%; or more preferably, between about 61% and about 100%; or even more preferably, between about 81% and about 100%; of nucleotides that are identical to the nucleotides of SEQ ID NO:1 or SEQ ID NO:3 will be sequences that are "essentially as set forth in SEQ ID NO:1 or SEQ ID NO:3". Sequences that are essentially the same as those set forth in SEQ ID NO:1 or SEQ ID NO:3 may also be functionally defined as sequences that are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:1 or SEQ ID NO:3 under relatively stringent conditions. Suitable relatively stringent hybridization conditions will be well known to those of skill in the art and are clearly set forth herein.

The present invention also encompasses DNA segments that are complementary, or essentially complementary, to the sequence set forth in SEQ ID NO:1 or SEQ ID NO:3. Nucleic acid sequences that are "complementary" are those that are capable of base-pairing according to the standard Watson-Crick complementarity rules. As used herein, the term "complementary sequences" means nucleic acid sequences that are substantially complementary, as may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:1 or SEQ ID NO:3 under relatively stringent conditions such as those described herein.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, nucleic acid fragments may be prepared that include a short contiguous stretch identical to or complementary to SEQ ID NO:1 or SEQ ID NO:3, such as about 14 nucleotides, and that

are up to about 10,000 or about 5,000 base pairs or about 3,000 base pairs in length, with segments of about 2,000 being preferred in certain cases. DNA segments with total lengths of about 1,000, about 500, about 200, about 100 and about 50 base pairs in length (including all intermediate lengths) are also contemplated to be useful.

It will be readily understood that "intermediate lengths", in these contexts, means any length between the quoted ranges, such as 14, 15, 16, 17, 18, 19, 20, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through the 200-500; 500-1,000; 1,000-2,000; 2,000-3,000; 3,000-5,000; 5,000-10,000 ranges, up to and including sequences of about 12,001, 12,002, 13,001, 13,002 and the like.

It will also be understood that this invention is not limited to the particular nucleic acid and amino acid sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4. Recombinant vectors and isolated DNA segments may therefore variously include the DPD coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include DPD-encoding regions or may encode biologically functional equivalent proteins or peptides that have variant amino acids sequences.

The DNA segments of the present invention encompass biologically functional equivalent DPD proteins and peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements to the antigenicity of the protein or to test DPD mutants in order to examine DPD activity at the molecular level.

If desired, one may also prepare fusion proteins and peptides, *e.g.*, where the DPD-encoding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes

(*e.g.*, proteins that may be purified by affinity chromatography and enzyme label coding regions, respectively).

Recombinant vectors form important further aspects of the present invention. Particularly useful vectors are contemplated to be those vectors in which the coding portion of the DNA segment, whether encoding a full length protein or smaller peptide, is positioned under the control of a promoter. The promoter may be in the form of the promoter that is naturally associated with DPD genes, *e.g.*, in mammalian cells, such as bovine or human cells, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCR™ technology, in connection with the compositions disclosed herein.

In other embodiments, it is contemplated that certain advantages will be gained by positioning the coding DNA segment under the control of a recombinant, or heterologous, promoter. As used herein, a recombinant or heterologous promoter is intended to refer to a promoter that is not normally associated with a DPD gene in its natural environment. Such promoters may include DPD promoters normally associated with other genes, and/or promoters isolated from any other bacterial, viral, eukaryotic, or mammalian cell. Naturally, it will be important to employ a promoter that effectively directs the expression of the DNA segment in the cell type, organism, or even animal, chosen for expression. The use of promoter and cell type combinations for protein expression is generally known to those of skill in the art of molecular biology, (for example, see Sambrook *et al.*, 1989). The promoters employed may be constitutive, or inducible, and can be used under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins or peptides. Appropriate promoter systems contemplated for use in high-level expression are well-known to those of skill in the art, and include such systems as the T7 RNA polymerase promoter system (Tabor and Richardson, 1985) and the maltose binding protein-fusion protein system (Guan *et al.*, 1987; Nagai and Thogersen, 1987).

As mentioned above, in connection with expression embodiments to prepare recombinant DPD proteins and peptides, it is contemplated that longer DNA segments will

most often be used, with DNA segments encoding the entire DPD protein or functional domains, subunits, *etc.* being most preferred. However, it will be appreciated that the use of shorter DNA segments to direct the expression of DPD peptides or epitopic core regions, such as may be used to generate anti-DPD antibodies, also falls within the scope of the invention.

DNA segments that encode peptide antigens from about 15 to about 50 amino acids in length, or more preferably, from about 15 to about 30 amino acids in length are contemplated to be particularly useful. The peptides may, of course, be of any length in this range, such as 16, 17, 18, 19, 20, or about 25 amino acids in length.

The DPD gene and DNA segments may also be used in connection with somatic expression in an animal or in the creation of a transgenic animal. Again, in such embodiments, the use of a recombinant vector that directs the expression of the full length or active DPD protein is particularly contemplated. The methods for preparation of transgenic animals and the transfer of DNA segments for expression in mammals are well known to those of skill in the art, as exemplified by U.S. Patent 4,396,601, incorporated herein by reference.

In addition to their use in directing the expression of the DPD protein, the nucleic acid sequences disclosed herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments.

3. Nucleic Acid Hybridization

In connection with expression embodiments to prepare recombinant DPD proteins and peptides, it is contemplated that longer DNA segments will most often be used, with DNA segments encoding DPD or the entire DPD protein being most preferred. However, it will be appreciated that the use of shorter DNA segments to direct the expression of DPD peptides or epitopic core regions, such as may be used to generate anti-DPD antibodies, also falls within the scope of the invention.

DNA segments that encode peptide antigens from about 14 to about 50 amino acids in length, or more preferably, from about 14 to about 30 amino acids in length are contemplated to be particularly useful, as are DNA segments encoding entire DPD proteins. The peptides may, of course, be of any length in this range, such as about 14,

15, 16, 17, 18, 19 or about 20 amino acids in length. This is the meaning of "about" in about 14, about 20, about 25, about 30, about 35, about 40, about 45 or about 50 amino acids in length, with "about", in this one context meaning a range of from 1 to 4 amino acids longer or shorter than the stated length, with 13 or 14 or so still being the minimum length. DNA segments encoding peptides will generally have a minimum coding length, or coding sequence region, in the order of about 45 to about 150, or to about 90 nucleotides. DNA segments encoding full-length proteins may have a minimum coding length in the order of about 3075 to about 4414 nucleotides for a protein in accordance with SEQ ID NO:2, or on the order of about 3075 to about 4368 nucleotides for a protein in accordance with SEQ ID NO:4.

In addition to their use in directing the expression of the DPD protein, the nucleic acid sequences disclosed herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments. As such, it is contemplated that nucleic acid segments that comprise a sequence region that consists of at least a 14 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 14 nucleotide long contiguous sequence of SEQ ID NO:1 or SEQ ID NO:3 will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000, 2000, 3000, 4000, *etc.*, (including all intermediate lengths) and even up to the full length sequence of about 4414 nucleotides in length for SEQ ID NO:1, and even up to the full length sequence of about 4368 nucleotides in length for SEQ ID NO:3, will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to DPD-encoding sequences will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are envisioned, including the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Nucleic acid molecules having sequence regions consisting of contiguous nucleotide stretches of about 10, 15, 30, 50, or even of about 100 to about 200 nucleotides or so, identical or complementary to SEQ ID NO:1 or SEQ ID NO:3 are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern

blotting. This would allow DPD structural or regulatory genes to be analyzed, both in eukaryotic and prokaryotic cells, including *e.g.,* mammalian cells such as human and bovine cells, or various bacterial or other prokaryotic species. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 14 and about 100 nucleotides, but larger contiguous complementary stretches of up to and including full-length nucleotides may be used, according to the length complementary sequences one wishes to detect.

The DNA segments of the present invention encompass biologically functional equivalent DPD proteins and peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.,* to introduce improvements to the antigenicity of the protein or to test DPD-deficient or DPD-altered mutants in order to examine DPD activity and FUra toxicity at the molecular level.

If desired, one may also prepare fusion proteins and peptides, *e.g.,* where the DPD coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes (*e.g.,* proteins that may be purified by affinity chromatography and enzyme label coding regions, respectively).

4. Recombinant Vectors Expressing DPD

A particular aspect of this invention provides novel ways in which to utilize DPD-encoding DNA segments and recombinant vectors comprising DPD-encoding DNA segments. As is well known to those of skill in the art, many such vectors are readily available, one particular detailed example of a suitable vector for expression in mammalian cells is that described in U. S. Patent 5,168,050, incorporated herein by reference. However, there is no requirement that a highly purified vector be used, so long as the coding segment employed encodes a DPD protein and does not include any coding or regulatory sequences that would have an adverse effect on cells. Therefore, it will also be understood that useful nucleic acid sequences may include additional residues, such as additional non-coding sequences flanking either of the 5′ or 3′ portions of the coding region or may include various internal sequences, *i.e.*, introns, which are known to occur within genes.

After identifying an appropriate DPD-encoding gene or DNA molecule, it may be inserted into any one of the many vectors currently known in the art, so that it will direct the expression and production of the DPD protein when incorporated into a host cell. In a recombinant expression vector, the coding portion of the DNA segment is positioned under the control of a promoter. The promoter may be in the form of the promoter which is naturally associated with a DPD-encoding gene, as may be obtained by isolating the 5′ non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCR™ technology, in connection with the compositions disclosed herein.

In certain embodiments, it is contemplated that particular advantages will be gained by positioning the DPD-encoding DNA segment under the control of a recombinant, or heterologous, promoter. As used herein, a recombinant or heterologous promoter is intended to refer to a promoter that is not normally associated with a DPD-encoding gene in its natural environment. Such promoters may include those normally associated with other DPD genes, and/or promoters isolated from any other bacterial, viral, eukaryotic, or mammalian cell. Naturally, it will be important to employ a promoter that effectively

directs the expression of the DNA segment in the particular cell containing the vector comprising the DPD gene.

The use of recombinant promoters to achieve protein expression is generally known to those of skill in the art of molecular biology, for example, see Sambrook *et al.*, (1989). The promoters employed may be constitutive, or inducible, and can be used under the appropriate conditions to direct high level or regulated expression of the introduced DNA segment. The currently preferred promoters are those such as CMV, RSV LTR, the SV40 promoter alone, and the SV40 promoter in combination with the SV40 enhancer.

5. DPD Pharmaceutical Compositions

Another aspect of the present invention includes novel compositions comprising isolated and purified DPD protein or nucleic acids which encode DPD protein. It will, of course, be understood that one or more than one DPD-encoding gene may be used in the methods and compositions of the invention. The nucleic acid delivery methods may thus entail the administration of one, two, three, or more, DPD-encoding genes. The maximum number of genes that may be applied is limited only by practical considerations, such as the effort involved in simultaneously preparing a large number of gene constructs or even the possibility of eliciting an adverse cytotoxic effect.

The particular combination of genes may be two or more distinct DPD-encoding genes; or it may be such that a DPD-encoding gene is combined with another gene and/or another protein such as a cytoskeletal protein, cofactor or other biomolecule; a hormone or growth factor gene may even be combined with a gene encoding a cell surface receptor capable of interacting with the polypeptide product of the first gene.

In using multiple genes, they may be combined on a single genetic construct under control of one or more promoters, or they may be prepared as separate constructs of the same or different types. Thus, an almost endless combination of different genes and genetic constructs may be employed. Certain gene combinations may be designed to, or their use may otherwise result in, achieving synergistic effects on cell growth and/or stimulation of an immune response. Any and all such combinations are intended to fall within the scope of the present invention. Indeed, many synergistic effects have been

described in the scientific literature, so that one of ordinary skill in the art would readily be able to identify likely synergistic gene combinations, or even gene-protein combinations.

It will also be understood that, if desired, the nucleic acid segment or gene encoding DPD could be administered in combination with further agents, such as, *e.g.*, proteins or polypeptides or various pharmaceutically active agents. So long as the composition comprises a DPD gene, there is virtually no limit to other components which may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The nucleic acids may thus be delivered along with various other agents as required in the particular instance.

Pharmaceutical compositions prepared in accordance with the present invention find use in preventing or ameliorating sepsis in an animal exposed to bacterial lipopolysaccharide. Such methods generally involve administering to an animal a pharmaceutical composition comprising an immunologically effective amount of a DPD composition. This composition may include an immunologically-effective amount of either a DPD peptide or a DPD-encoding nucleic acid composition. Such compositions may also be used to generate an immune response in an animal.

Therapeutic kits comprising DPD peptides or DPD-encoding nucleic acid segments comprise another aspect of the present invention. Such kits will generally contain, in suitable container, a pharmaceutically acceptable formulation of DPD peptide or a DPD-encoding nucleic acid composition. The kit may have a single container that contains the DPD composition or it may have distinct containers for the DPD composition and other reagents which may be included within such kits.

The components of the kit may be provided as liquid solution(s), or as dried powder(s). When the components are provided in a liquid solution, the liquid solution is an aqueous solution, with a sterile aqueous solution being particularly preferred. When reagents or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent. It is envisioned that the solvent may also be provided in another container.

In related embodiments, the present invention contemplates the preparation of diagnostic kits that may be employed to detect the presence of DPD proteins or peptides

and/or antibodies in a sample. Generally speaking, kits in accordance with the present invention will include a suitable DPD protein or peptide or antibody directed against such a protein or peptide, together with an immunodetection reagent and a container for the antibody or antigen and reagent. The components of the diagnostic kits may be packaged either in aqueous media or in lyophilized form.

The immunodetection reagent will typically comprise a label associated with the antibody or antigen, or associated with a secondary binding ligand. Exemplary ligands might include a secondary antibody directed against the first antibody or antigen or a biotin or avidin (or streptavidin) ligand having an associated label. Of course, as noted above, a number of exemplary labels are known in the art and all such labels may be employed in connection with the present invention. The kits may contain antibody-label conjugates either in fully conjugated form, in the form of intermediates, or as separate moieties to be conjugated by the user of the kit.

The container will generally include at least one vial, test tube, flask, bottle, syringe or other container, into which the antigen or antibody may be placed, and preferably suitably aliquoted. Where a second binding ligand is provided, the kit will also generally contain a second vial or other container into which this ligand or antibody may be placed. The kits of the present invention will also typically include a container for the antibody, antigen, and reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained.

6. Methods of DNA Transfection

Technology for introduction of DNA into cells is well-known to those of skill in the art. Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham and van der Eb, 1973; Zatloukal *et al.*, 1992); (2) physical methods such as microinjection (Capecchi, 1980), electroporation (Wong and Neumann, 1982; Fromm *et al.*, 1985) and the gene gun (Johnston and Tang, 1994; Fynan *et al.*, 1993); (3) viral vectors (Clapp, 1993; Lu *et al.*, 1993; Eglitis and Anderson, 1988a; 1988b); and (4) receptor-mediated mechanisms (Curiel *et al.*, 1991; 1992; Wagner *et al.*, 1992).

7. Liposomes and Nanocapsules

The formation and use of liposomes is generally known to those of skill in the art (see for example, Couvreur *et al.*, 1988 which describes the use of liposomes and nanocapsules in the targeted antibiotic therapy of intracellular bacterial infections and diseases). Recently, liposomes were developed with improved serum stability and circulation half-times (Gabizon and Papahadjopoulos, 1988; Allen and Choun, 1987; Desiderio and Campbell, 1983).

Nanocapsules can generally entrap compounds in a stable and reproducible way (Henry-Michelland *et al.*, 1987). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) should be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present invention, and such particles may be are easily made, as described (Couvreur *et al.*, 1977; 1988).

Liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs). MLVs generally have diameters of from 25 nm to 4 μm. Sonication of MLVs results in the formation of small unilamellar vesicles (SUVs) with diameters in the range of 200 to 500 Å, containing an aqueous solution in the core.

In addition to the teachings of Couvreur *et al.* (1988), the following information may be utilized in generating liposomal formulations. Phospholipids can form a variety of structures other than liposomes when dispersed in water, depending on the molar ratio of lipid to water. At low ratios the liposome is the preferred structure. The physical characteristics of liposomes depend on pH, ionic strength and the presence of divalent cations. Liposomes can show low permeability to ionic and polar substances, but at elevated temperatures undergo a phase transition which markedly alters their permeability. The phase transition involves a change from a closely packed, ordered structure, known as the gel state, to a loosely packed, less-ordered structure, known as the fluid state. This occurs at a characteristic phase-transition temperature and results in an increase in permeability to ions, sugars and drugs.

Liposomes interact with cells via four different mechanisms: Endocytosis by phagocytic cells of the reticuloendothelial system such as macrophages and neutrophils; adsorption to the cell surface, either by nonspecific weak hydrophobic or electrostatic forces, or by specific interactions with cell-surface components; fusion with the plasma cell membrane by insertion of the lipid bilayer of the liposome into the plasma membrane, with simultaneous release of liposomal contents into the cytoplasm; and by transfer of liposomal lipids to cellular or subcellular membranes, or *vice versa*, without any association of the liposome contents. It often is difficult to determine which mechanism is operative and more than one may operate at the same time.

8. DPD Compositions and FUra Treatment of Proliferative Cell Disorders

Another aspect of the present invention is the use of DPD in the treatment of proliferative cell disorders when the cancer therapeutic drug FUra is indicated as a treatment of such disorders. In patients who have limited DPD expression, or in patients where a DPD deficiency is noted, it may be desirable to concomitantly administer pharmaceutically-acceptable compositions of DPD. Such compositions may include DPD protein or DPD-encoding DNA segments in accordance with the present invention.

9. Expression of DPD

For the expression of DPD, once a suitable (full-length if desired) clone or clones have been obtained, whether they be cDNA based or genomic, one may proceed to prepare an expression system for the recombinant preparation of DPD. The engineering of DNA segment(s) for expression in a prokaryotic or eukaryotic system may be performed by techniques generally known to those of skill in recombinant expression. It is believed that virtually any expression system may be employed in the expression of DPD.

DPD may be successfully expressed in eukaryotic expression systems, however, it is also envisioned that bacterial expression systems may be preferred for the preparation of DPD for all purposes. The cDNA for DPD may be separately expressed in bacterial systems, with the encoded proteins being expressed as fusions with β-galactosidase, ubiquitin, *Schistosoma japonicum* glutathione S-transferase, and the like. It is believed

that bacterial expression will ultimately have advantages over eukaryotic expression in terms of ease of use and quantity of materials obtained thereby.

It is proposed that transformation of host cells with DNA segments encoding DPD will provide a convenient means for obtaining DPD peptide. Both cDNA and genomic sequences are suitable for eukaryotic expression, as the host cell will, of course, process the genomic transcripts to yield functional mRNA for translation into protein.

It is similarly believed that almost any eukaryotic expression system may be utilized for the expression of DPD, *e.g.,* baculovirus-based, glutamine synthase-based or dihydrofolate reductase-based systems could be employed. However, in preferred embodiments, it is contemplated that plasmid vectors incorporating an origin of replication and an efficient eukaryotic promoter, as exemplified by the eukaryotic vectors of the pCMV series, such as pCMV5, will be of most use.

For expression in this manner, one would position the coding sequences adjacent to and under the control of the promoter. It is understood in the art that to bring a coding sequence under the control of such a promoter, one positions the 5' end of the transcription initiation site of the transcriptional reading frame of the protein between about 1 and about 50 nucleotides "downstream" of (*i.e.,* 3' of) the chosen promoter.

Where eukaryotic expression is contemplated, one will also typically desire to incorporate into the transcriptional unit which includes DPD, an appropriate polyadenylation site (*e.g.,* 5'-AATAAA-3') if one was not contained within the original cloned segment. Typically, the poly A addition site is placed about 30 to 2000 nucleotides "downstream" of the termination site of the protein at a position prior to transcription termination.

It is contemplated that virtually any of the commonly employed host cells can be used in connection with the expression of DPD in accordance herewith. Examples include cell lines typically employed for eukaryotic expression such as 239, AtT-20, HepG2, VERO, HeLa, CHO, WI 38, BHK, COS-7, RIN and MDCK cell lines.

It is contemplated that DPD may be "overexpressed", *i.e.,* expressed in increased levels relative to its natural expression in human cells, or even relative to the expression of other proteins in a recombinant host cell containing DPD-encoding DNA segments.

Such overexpression may be assessed by a variety of methods, including radio-labeling and/or protein purification. However, simple and direct methods are preferred, for example, those involving SDS/PAGE and protein staining or Western blotting, followed by quantitative analyses, such as densitometric scanning of the resultant gel or blot. A specific increase in the level of the recombinant protein or peptide in comparison to the level in natural DPD-producing animal cells is indicative of overexpression, as is a relative abundance of the specific protein in relation to the other proteins produced by the host cell and, *e.g.,* visible on a gel.

As used herein, the term "engineered" or "recombinant" cell is intended to refer to a cell into which a recombinant gene, such as a gene encoding a DPD peptide has been introduced. Therefore, engineered cells are distinguishable from naturally occurring cells which do not contain a recombinantly introduced gene. Engineered cells are thus cells having a gene or genes introduced through the hand of man. Recombinantly introduced genes will either be in the form of a cDNA gene (*i.e.,* they will not contain introns), a copy of a genomic gene, or will include genes positioned adjacent to a promoter not naturally associated with the particular introduced gene.

Generally speaking, it may be more convenient to employ as the recombinant gene a cDNA version of the gene. It is believed that the use of a cDNA version will provide advantages in that the size of the gene will generally be much smaller and more readily employed to transfect the targeted cell than will a genomic gene, which will typically be up to an order of magnitude larger than the cDNA gene. However, the inventors do not exclude the possibility of employing a genomic version of a particular gene where desired.

Where the introduction of a recombinant version of one or more of the foregoing genes is required, it will be important to introduce the gene such that it is under the control of a promoter that effectively directs the expression of the gene in the cell type chosen for engineering. In general, one will desire to employ a promoter that allows constitutive (constant) expression of the gene of interest. Commonly used constitutive promoters are generally viral in origin, and include the cytomegalovirus (CMV) promoter, the Rous sarcoma long-terminal repeat (LTR) sequence, and the SV40 early gene promoter. The use of these constitutive promoters will ensure a high, constant level of

expression of the introduced genes. The inventors have noticed that the level of expression from the introduced genes of interest can vary in different clones, probably as a function of the site of insertion of the recombinant gene in the chromosomal DNA. Thus, the level of expression of a particular recombinant gene can be chosen by evaluating different clones derived from each transfection experiment; once that line is chosen, the constitutive promoter ensures that the desired level of expression is permanently maintained. It may also be possible to use promoters that are specific for cell type used for engineering, such as the insulin promoter in insulinoma cell lines, or the prolactin or growth hormone promoters in anterior pituitary cell lines.

10. Enhanced Production of Bovine and Human DPD

DNA segments of the present invention and the novel methods for isolation of active human DPD provide significant improvements over the limited successes of isolating large quantities of native, active DPD from such natural sources as bovine or human cells. The novel purification processes disclosed herein permit the facile isolation of large quantities of the human and bovine proteins, and in combination with recombinant DNA methodologies well-known to those of skill in the art, permit the rapid isolation of large quantities of recombinant proteins.

An aspect of the present invention is the enhanced production of DPD by recombinant methodologies in a bacterial host, employing DNA constructs to transform Gram-positive or Gram-negative bacterial cells. For example, the use of *Escherichia coli* expression systems are well known to those of skill in the art, as is the use of other bacterial species such as *Bacillus subtilis* or *Streptococcus sanguis*.

Further aspects of the invention include high expression vectors incorporating DNA encoding the novel DPD and its variants. It is contemplated that vectors providing enhanced expression of DPD in other systems such as *S. mutans* will also be obtainable. Where it is desirable, modifications of the physical properties of DPD may be sought to increase its solubility or expression in liquid culture. The DPD-encoding locus may be placed under control of a high expression promoter or the components of the expression system altered to enhance expression.

11. DPD Antibodies

In another aspect, the present invention contemplates an antibody that is immunoreactive with a polypeptide of the invention. An antibody can be a polyclonal or a monoclonal antibody. In a preferred embodiment, an antibody is a monoclonal antibody. Means for preparing and characterizing antibodies are well known in the art (See, *e.g.*, Howell and Lane, 1988).

Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogen comprising a polypeptide of the present invention and collecting antisera from that immunized animal. A wide range of animal species can be used for the production of antisera. Typically an animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster or a guinea pig. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

Antibodies, both polyclonal and monoclonal, specific for DPD may be prepared using conventional immunization techniques, as will be generally known to those of skill in the art. A composition containing antigenic epitopes of DPD can be used to immunize one or more experimental animals, such as a rabbit or mouse, which will then proceed to produce specific antibodies against DPD. Polyclonal antisera may be obtained, after allowing time for antibody generation, simply by bleeding the animal and preparing serum samples from the whole blood.

To obtain monoclonal antibodies, one would also initially immunize an experimental animal, often preferably a mouse, with a DPD composition. One would then, after a period of time sufficient to allow antibody generation, obtain a population of spleen or lymph cells from the animal. The spleen or lymph cells can then be fused with cell lines, such as human or mouse myeloma strains, to produce antibody-secreting hybridomas. These hybridomas may be isolated to obtain individual clones which can then be screened for production of antibody to the desired DPD peptide.

Following immunization, spleen cells are removed and fused, using a standard fusion protocol with plasmacytoma cells to produce hybridomas secreting monoclonal antibodies against DPD. Hybridomas which produce monoclonal antibodies to the selected antigens are identified using standard techniques, such as ELISA and Western blot

methods. Hybridoma clones can then be cultured in liquid media and the culture supernatants purified to provide the DPD-specific monoclonal antibodies.

It is proposed that the monoclonal antibodies of the present invention will find useful application in standard immunochemical procedures, such as ELISA and Western blot methods, as well as other procedures which may utilize antibody specific to DPD epitopes.

Additionally, it is proposed that monoclonal antibodies specific to DPD may be utilized in other useful applications. For example, their use in immunoabsorbent protocols may be useful in purifying native or recombinant DPD species or variants thereof.

In general, both poly- and monoclonal antibodies against DPD may be used in a variety of embodiments. For example, they may be employed in antibody cloning protocols to obtain cDNAs or genes encoding DPD or related proteins. They may also be used in inhibition studies to analyze the effects of DPD in cells or animals. Anti-DPD antibodies will also be useful in immunolocalization studies to analyze the distribution of DPD during various cellular events, for example, to determine the cellular or tissue-specific distribution of the DPD peptide under different physiological conditions. A particularly useful application of such antibodies is in purifying native or recombinant DPD, for example, using an antibody affinity column. The operation of all such immunological techniques will be known to those of skill in the art in light of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

The drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1A. Cloning Strategy of Bovine Liver DPD cDNA. A degenerate sense primer (designated "Primer A") and an antisense primer (designated "Primer B") were designed from oligonucleotides based on the sequence of a 23 amino acid tryptic fragment (KAEASGAXALELNLSCPHGMGER; SEQ ID NO:7) obtained from purified bovine liver DPD. The primers correspond to the respective amino acid peptide sequences KAEASGA (SEQ

ID NO:10) and PHGMGER (SEQ ID NO:11) (denoted by boxes). The 65 base pair amplified DPD cDNA was subsequently subcloned and sequenced. The region between Primer A and Primer B found to encode the predicted amino acids XALELNLSC (SEQ ID NO:24). The previously unknown amino acid (X) was determined to be Asp (D).

FIG. 1B. Based on specific sequence obtained from the 65 base pair fragment, primer C coding for DALELNLSC (SEQ ID NO:5) together with the antisense primer specific to the adaptor region of oligo(dT) was used to PCR™ amplify a 2360 base pair product corresponding to the 3' end of the DPD clone.

FIG. 1C. The 5' end of the DPD cDNA was amplified using degenerate primer D designed to the N-terminal amino acid sequence of purified bovine liver DPD (KDVADIE) (SEQ ID NO:6) along with primer E corresponding to sequence obtained from the 2360 base pair fragment (FIG. 1B).

FIG. 1D. The initiating ATG was obtained by using primer F derived from sequence obtained from the 2076 base pair fragment (FIG. 1C) along with anchor primer to generate a 237 base pair PCR™ product.

FIG. 2. Alignment and Ligation of the Full-Length Bovine Liver DPD cDNA. The full-length clone (4414 base pairs; SEQ ID NO:1) was generated by ligation of the three cDNA fragments (2360, 2076, and 237 base pairs). Each fragment was independently identified as a portion of the DPD clone by identification of specific peptide sequence derived from purified bovine liver DPD (as shown in boldface). Restriction sites common in only the overlapping regions (shown by hashed lines) were utilized to ligate the three fragments together.

FIG. 3. Northern Analysis of Bovine Liver DPD mRNA. The left lane contains 30 μg total RNA; the right lane contains 1 μg Poly(A)+ RNA. Samples were resolved on a 1.5% agarose-formaldehyde gel, transferred to nylon membrane, probed with ^{32}P-labeled DPD cDNA, and autoradiographed for 24 h at -80°C.

FIG. 4. *In vitro* translation of bovine liver cDNA. RNA from *in vitro* transcription of full-length bovine liver cDNA was translated using a rabbit reticulocyte lysate system. Translated protein was resolved by SDS-PAGE, transferred to nitrocellulose, and exposed to autoradiographic film. Lane 1 contains reaction products produced using empty vector

(pCRII®) DNA. Lane 2 contains reaction products produced by bovine liver cDNA demonstrating a 108 kDa band. Lane 3 contains reaction products produced by Luciferase positive control DNA demonstrating a 61 kDa band. Lane 4 contains reaction products produced in the absence of DNA (negative control).

FIG. 5. Immunoblot analysis of recombinant and bovine liver DPD. Proteins were resolved by SDS-PAGE, electrophoretically transferred to nitrocellulose, and reacted with a 1:20,000-fold dilution of the rabbit anti-DPD polyclonal antibody. The prestained standards and their corresponding molecular weights are labeled. The left lane contains 0.5 μg purified bovine liver DPD. The center lane contains cytosol from induced cells eluted from the amylose column after factor Xa cleavage. The right lane contains fusion protein purified from the amylose column prior to cleavage with factor Xa.

FIG. 6A. Nucleotide and predicted amino acid sequence of bovine liver DPD. This figure is presented on three panels, FIG. 6A, FIG. 6B and FIG. 6C. The amino acid sequences of peptides derived from either N-terminal sequencing (4-13), tryptic digestion (656-678) or CNBr digestion (743-760) of purified bovine liver DPD are underlined. The bovine DPD amino acid sequence was compared to other sequences in the computer data base. Regions of similarity occurring within the flavin binding domain of dihydroorotate dehydrogenase (amino acid residues 788-795) and the flavin-NADPH binding domains of thioredoxin reductase (residues 187-204 and 332-348, respectively) are indicated on the bovine DPD sequence by overlining. Both the initiating (ATG) and stop (TAA) codon are indicated in boldface type. PCR™ primers (indicated by bold face and underlining) used in the amplification of bovine liver cDNA are as follows: primers A, B and C contained within nucleotides 2040-2108; primer D at nucleotides 93-112; primer E at nucleotides 2149-2168; primer F at 218-237.

FIG. 6B. Nucleotide and predicted amino acid sequence of bovine liver DPD. This figure is panel two of three.

FIG. 6C. Nucleotide and predicted amino acid sequence of bovine liver DPD. This figure is panel three of three.

FIG. 7A. Nucleotide and predicted amino acid sequence of human lymphocyte DPD. This figure is presented on four panels, FIG. 7A, FIG. 7B, FIG. 7C, and FIG. 7D.

Both the initiating (ATG) and stop (TAA) codon are indicated in boldface type. For amplification of human lymphocyte DPD cDNA from a normal (normal DPD activity) and DPD deficient patient, the following primers were used: (sense 5'-TGTAGGCACTGCCATGGCCCCTGTG-3') (SEQ ID NO:25) and (antisense 5'-TTCACAAATCACCTTAACACACC-3') (SEQ ID NO:26). These primers correspond to positions 36-60 and 3117-3139, respectively, of the DPD cDNA sequence (boxed). For amplification from genomic DNA, the primers used were: (sense 5'-TTGGTGGTTTAAGTACTTCTGAAATTCC-3' (SEQ ID NO:27) and antisense 5'-CTTGCTCTGTCCGAACAAACTGCATAGCA-3' (SEQ ID NO:28), corresponding to positions 716-743 and 1260-1288, respectively (boxed). The single amino acid difference between human lymphocyte (N) and human liver (S) DPD is shown at amino acid position 910 (arrow). The single nucleotide deletion resulting in a frameshift in the DPD deficient patient is shown (arrow) at nucleotide position 1000 (corresponding to codon 318).

FIG. 7B. Nucleotide and predicted amino acid sequence of human lymphocyte DPD. This figure is panel two of four panels.

FIG. 7C. Nucleotide and predicted amino acid sequence of human lymphocyte DPD. This figure is panel three of four panels, FIG. 7A, FIG. 7B, FIG. 7C, and FIG. 7D.

FIG. 7D. Nucleotide and predicted amino acid sequence of human lymphocyte DPD. This figure is panel four of four panels, FIG. 7A, FIG. 7B, FIG. 7C, and FIG. 7D.

FIG. 8. Western blot analysis of cytosol from the DPD deficient patient and the normal subject. Lanes 1 and 2 contain 100 and 200 μg lymphocyte cytosol, respectively, from the DPD-deficient patient. Lane 3 contains 50 μg cytosol from the normal control (normal DPD activity).

FIG. 9. Northern blot analysis of total and poly(A)+ RNA from human lymphocytes from normal and DPD-deficient individuals. Lanes 1 and 3 contain 30 μg of total human lymphocyte RNA from the normal and DPD-deficient individual, respectively. Lanes 2 and 4 contain 2 μg poly(A)+ RNA from the normal and DPD-deficient individual, respectively.

FIG. 10A. *In vitro* transcription/translation of human DPD cDNA. RNA from *in vitro* transcription of human lymphocyte DPD cDNA was translated in the presence of

[^{35}S] methionine for labeling of the synthesized proteins; lane 1 contains reaction products produced by human lymphocyte cDNA cloned from an individual with normal DPD enzyme activity and demonstrates a 108,000 dalton band. Lane 2 contains reaction products produced using the cDNA cloned from the DPD-deficient patient and demonstrates a 40,000 dalton band corresponding to truncated DPD. Lane 3 contains a luciferase positive control demonstrating a 61,000 dalton band.

FIG. 10B. *In vitro* transcription/translation of human DPD cDNA. RNA from *in vitro* transcription of human lymphocyte DPD cDNA was translated in the presence of unlabeled amino acid for western blot analysis; lane 4 contains prestained molecular weight markers. Lane 5 contains 0.2 μg purified DPD (Lu *et al.*, 1992). Lane 6 contains reaction products produced by human lymphocyte cDNA cloned from an individual with normal DPD enzyme activity and demonstrates a 108,000 dalton band corresponding to the band seen in lane 1. Lane 7 contains reaction products produced by the cDNA cloned from the DPD deficient patient and demonstrates a 40,000 dalton band corresponding to the band seen in lane 2. Lane 8 contains reaction products produced by luciferase positive control.

FIG. 11. Partial nucleotide sequence demonstrating the adenosine deletion in the genomic DNA of the DPD-deficient patient. Sequence analysis of the DPD deficient patient's genomic DNA revealed a single adenosine deletion as compared to the individual with normal DPD activity. This deletion was identified in both the cDNA and the genomic DNA and causes a frameshift in codon 318.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Based on the amino acid sequence of peptides derived from purified bovine liver DPD, the full-length cDNA was cloned, sequenced, and expressed in a bacterial cell line. Comparison to other sequences in the GenBank data base verified that this is a unique sequence. The human liver DPD gene was shown to encode a polypeptide of 1025 amino acids (M_r = 111,688 daltons) which corresponded to a monomer of purified dimeric enzyme (Lu *et al.*, 1993). Northern blot analysis of bovine liver RNA detected a single band of appropriate length corresponding to the full-length cDNA, and bacterial expression

of the DPD cDNA generated a protein which comigrated with purified bovine liver DPD during SDS-PAGE. This peptide, when immunoblotted, also reacted with a specific polyclonal rabbit anti-DPD antibody. Analysis of the bovine liver cDNA indicated the presence of FAD, NADPH and 4-Fe/4-S binding site prosthetic groups within the translated polypeptide.

A partial DPD cDNA (2300 base pairs long) was isolated from a λgt11 bovine liver cDNA library (Clontech) and found to have 100% sequence identity with the full-length bovine cDNA identified. The partial clone began at nucleotide 425 and extended through to nucleotide 2765. Examination of the sequence following this region (nucleotides 2766-2776 of SEQ ID NO:1) indicated a short poly-A tract which may have served as a start site for first-strand cDNA synthesis during construction of the library. This DPD cDNA fragment was isolated from the bovine liver cDNA library as a single clone and codes for most of the open reading frame of bovine liver cDNA. When translated, the amino acid sequence includes both the CNBr and tryptic peptide sequences derived from purified bovine liver DPD. Isolation of this partial DPD cDNA provides further evidence that this newly described full-length cDNA, which is a composite of three separate cDNA fragments, encodes DPD.

Data base searches for amino acid sequences, similar to DPD, identified dihydroorotate dehydrogenase, thioredoxin reductase, and glutamate synthase with a partial amino acid sequence identity of 40, 37, and 38%, respectively. While these values are too low to support a common ancestry for these proteins (Doolittle, 1981), they do contain certain functional similarities to DPD. Dihydroorotate dehydrogenase is a flavoprotein (using FAD as a cofactor) which catalyzes the fourth step in pyrimidine biosynthesis (Quinn *et al.*, 1991). In addition, both thioredoxin reductase and glutamate synthase use NADPH as a cofactor (Russel and Model, 1988; Oliver *et al.*, 1987). Alignment analyses did not identify unique and distinct FAD and NADPH binding sites in bovine liver cDNA, because of the close proximity and sharing of common elements of these two regions. Further analysis of the translated bovine DPD cDNA sequence revealed one 4-Fe/4-S binding site, a GDP/GTP binding site and a cAMP and cGMP

dependent protein kinase phosphorylation site. The iron sulfur binding site is consistent with data obtained from purified DPD (Shiotani and Weber, 1981; Lu *et al.*, 1992).

2. Immunoassays

As noted, it is proposed that the human DPD polypeptide of the invention will find utility as immunogens, *e.g.*, in connection with vaccine development, or as antigens in immunoassays for the detection of anti-DPD antigen-reactive antibodies. Turning first to immunoassays, in their most simple and direct sense, preferred immunoassays of the invention include the various types of enzyme linked immunosorbent assays (ELISAs) known to the art. However, it will be readily appreciated that the utility of DPD peptides is not limited to such assays, and that other useful embodiments include RIAs and other non-enzyme linked antibody binding assays or procedures.

In one such ELISA, peptides incorporating the DPD antigen sequences of invention may be first immobilized onto a selected surface, *e.g.*, a well of a surface exhibiting a protein affinity, such as a well in a polystyrene microtiter plate. In such an ELISA, generally, labeled anti-DPD antibodies would then be added to the wells, allowed to bind, and detected by means of their label. The amount of DPD in an unknown sample would be determined by mixing the sample with the labeled anti-DPD antibodies before or during incubation with the DPD in the wells. The presence of DPD in the sample acts to reduce the amount of anti-DPD antibody available for binding to the well and thus reduces the ultimate signal.

In another form of ELISA, an antibody capable of binding a DPD protein or peptide of the invention may be immobilized onto the solid surface, or well, and used directly in conjunction with labeled DPD compositions. In these ELISAs, generally, labeled DPD is added to the wells, allowed to bind, and detected by means of the label. The amount of DPD in an unknown sample is here determined by mixing the sample with the labeled DPD before or during incubation with the anti-DPD antibody in the wells. The presence of DPD in the sample again acts to reduce the amount of labeled DPD available for binding to the well and thus reduces the ultimate signal.

In coating a plate with either antigen or antibody, one will generally wash the wells of the plate to remove incompletely adsorbed material and then bind or "coat" a

nonspecific protein onto the wells of the plate. Nonspecific proteins are those that are known to be antigenically neutral with regard to the test antisera, and include bovine serum albumin (BSA), casein and solutions of milk powder. The coating allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific binding of antisera onto the surface.

Where an antibody capable of binding a DPD polypeptide is immobilized onto an ELISA plate, it is more customary to use a secondary or tertiary detection means rather than a direct procedure. Thus, after binding of antibody to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the control DPD and/or clinical or biological sample to be tested in a manner conducive to immune complex (antigen/antibody) formation. Detection of the DPD then requires a labeled secondary antibody, or a secondary antibody and a labeled tertiary, antibody. The labeled secondary antibody is, of course, an anti-DPD antibody that is conjugated to a detectable label. When using a tertiary approach, the secondary antibody is an unlabeled anti-DPD antibody and the tertiary antibody is a labeled antibody that is specific for the species, or isotype, of the secondary antibody employed.

A "manner conducive to immune complex (antigen/antibody) formation" means that the conditions preferably include diluting the antigens and antibodies with solutions such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween®. These added agents also tend to assist in the reduction of nonspecific background.

Incubation steps are typically from about 1 to 2 to about 4 hours, at temperatures preferably on the order of about 25° to 27°C, or may be overnight at about 4°C or so. Following all incubation steps in an ELISA, the contacted surface is generally washed so as to remove non-immunocomplexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween©, or borate buffer.

Following the formation of specific immunocomplexes between the test sample and the originally bound material, and subsequent washing, the occurrence of even minute amounts of immunocomplexes may be determined. As mentioned above, this may be achieved by subjecting the first immunocomplex to a second antibody having specificity

for the first, or even a third antibody having specificity for the second. Where a second antibody alone is used, given that the control and test Fas samples will typically be of human origin, the second antibody will preferably be an antibody having specificity in general for human Fas. Where a third antibody is also used, the second antibody will still preferably be an antibody having specificity for human Fas, and the third antibody will then be an antibody having specificity in general for the second antibody. A second murine antibody and a third anti-mouse Ig antibody is a particular example.

To provide a detecting means, the second or third antibody will have an associated label to allow detection. Preferably, this will be an enzyme that will generate color development upon incubating with an appropriate chromogenic substrate. Thus, for example, one will desire to contact and incubate the first or second immunocomplex with a urease, glucose oxidase or peroxidase-conjugated antibody for a period of time and under conditions that favor the development of further immunocomplex formation (*e.g.*, incubation for 2 hours at room temperature in a PBS-containing solution such as PBS/Tween©).

After incubation with the labeled antibody, and subsequent to washing to remove unbound material, the amount of label is quantified, *e.g.*, by incubation with a chromogenic substrate such as urea and bromocresol purple or 2,2'-azino-di-(3-ethyl-benzthiazoline-6-sulfonic acid) [ABTS] and H_2O_2, in the case of peroxidase as the enzyme label. Quantitation is then achieved by measuring the degree of color generation, *e.g.*, using a visible spectrum spectrophotometer.

3. Nucleic Acid Embodiments

The use of a hybridization probe of about 10-14 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 10 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 14 to 20 contiguous nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 and to select any continuous portion of the sequence, from about 14 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a probe or primer. The choice of probe and primer sequences may be governed by various factors, such as, by way of example only, one may wish to employ primers from towards the termini of the total sequence, or from the ends of the functional domain-encoding sequences, in order to amplify further DNA; one may employ probes corresponding to the entire DNA, or to the carboxyterminal or aminoterminal region, to clone DPD-type genes from other species or to clone further DPD-like or homologous genes from any species including human; and one may employ wild-type and mutant probes or primers with sequences centered around the DPD sequence to screen DNA samples for DPD, such as to identify human subjects that carry the DPD deletion mutation and thus may be susceptible to DPD deficiency and FUra toxicity.

The process of selecting and preparing a nucleic acid segment that includes a contiguous sequence from within SEQ ID NO:1 or SEQ ID NO:3 may alternatively be described as preparing a nucleic acid fragment. Of course, fragments may also be obtained by other techniques such as, *e.g.*, by mechanical shearing or by restriction enzyme digestion. Small nucleic acid segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™ technology of U.S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicators are known in the art,

including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmental undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to specific hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, *etc.*). Following washing of the hybridized surface so as to remove nonspecifically bound probe molecules, specific hybridization is detected, or even quantified, by means of the label.

4. Epitopic Core Sequences

The present invention is also directed to protein or peptide compositions, free from total cells and other peptides, which comprise a purified protein or peptide which incorporates an epitope that is immunologically cross-reactive with one or more anti-DPD antibodies.

As used herein, the term "incorporating an epitope(s) that is immunologically cross-reactive with one or more anti-DPD antibodies" is intended to refer to a peptide or protein antigen which includes a primary, secondary or tertiary structure similar to an epitope located within a DPD polypeptide. The level of similarity will generally be to such a degree that monoclonal or polyclonal antibodies directed against the DPD polypeptide will also bind to, react with, or otherwise recognize, the cross-reactive peptide or protein antigen. Various immunoassay methods may be employed in

conjunction with such antibodies, such as, for example, Western blotting, ELISA, RIA, and the like, all of which are known to those of skill in the art.

The identification of DPD epitopes, and/or their functional equivalents, suitable for use in vaccines is a relatively straightforward matter. For example, one may employ the methods of Hopp, as taught in U.S. Patent 4,554,101, incorporated herein by reference, which teaches the identification and preparation of epitopes from amino acid sequences on the basis of hydrophilicity. The methods described in several other papers, and software programs based thereon, can also be used to identify epitopic core sequences (see, for example, Jameson and Wolf, 1988; Wolf *et al.*, 1988; U.S. Patent Number 4,554,101). The amino acid sequence of these "epitopic core sequences" may then be readily incorporated into peptides, either through the application of peptide synthesis or recombinant technology.

Preferred peptides for use in accordance with the present invention will generally be on the order of 10 to 50 amino acids in length, and more preferably about 20 to about 40 amino acids in length. It is proposed that shorter antigenic DPD peptides will provide advantages in certain circumstances, for example, in the preparation of vaccines or in immunologic detection assays. Exemplary advantages include the ease of preparation and purification, the relatively low cost and improved reproducibility of production, and advantageous biodistribution.

It is proposed that particular advantages of the present invention may be realized through the preparation of synthetic peptides which include modified and/or extended epitopic/immunogenic core sequences which result in a "universal" epitopic peptide directed to DPD sequences. These epitopic core sequences are identified herein in particular aspects as hydrophilic regions of the DPD polypeptide antigen. It is proposed that these regions represent those which are most likely to promote T-cell or B-cell stimulation, and, hence, elicit specific antibody production.

An epitopic core sequence, as used herein, is a relatively short stretch of amino acids that is "complementary" to, and therefore will bind, antigen binding sites on transferrin-binding protein antibodies. Additionally or alternatively, an epitopic core sequence is one that will elicit antibodies that are cross-reactive with antibodies directed

against the peptide compositions of the present invention. It will be understood that in the context of the present disclosure, the term "complementary" refers to amino acids or peptides that exhibit an attractive force towards each other. Thus, certain epitope core sequences of the present invention may be operationally defined in terms of their ability to compete with or perhaps displace the binding of the desired protein antigen with the corresponding protein-directed antisera.

In general, the size of the polypeptide antigen is not believed to be particularly crucial, so long as it is at least large enough to carry the identified core sequence or sequences. The smallest useful core sequence anticipated by the present disclosure would generally be on the order of about 8 to about 10 amino acids in length, with sequences on the order of 15 to 25 being more preferred. Thus, this size will generally correspond to the smallest peptide antigens prepared in accordance with the invention. However, the size of the antigen may be larger where desired, so long as it contains a basic epitopic core sequence.

The identification of epitopic core sequences is known to those of skill in the art, for example, as described in U.S. Patent 4,554,101, incorporated herein by reference, which teaches the identification and preparation of epitopes from amino acid sequences on the basis of hydrophilicity. Moreover, numerous computer programs are available for use in predicting antigenic portions of proteins (see *e.g.*, Jameson and Wolf, 1988; Wolf *et al.*, 1988). Computerized peptide sequence analysis programs (*e.g.*, DNAStar™, DNAStar, Inc., Madison, WI) may also be useful in designing synthetic peptides in accordance with the present disclosure.

Syntheses of epitopic sequences, or peptides which include an antigenic epitope within their sequence, are readily achieved using conventional synthetic techniques such as the solid phase method (*e.g.*, through the use of commercially available peptide synthesizer such as an Applied Biosystems Model 430A Peptide Synthesizer). Peptide antigens synthesized in this manner may then be aliquotted in predetermined amounts and stored in conventional manners, such as in aqueous solutions or, even more preferably, in a powder or lyophilized state pending use.

In general, due to the relative stability of peptides, they may be readily stored in aqueous solutions for fairly long periods of time if desired, *e.g.*, up to six months or more, in virtually any aqueous solution without appreciable degradation or loss of antigenic activity. However, where extended aqueous storage is contemplated it will generally be desirable to include agents including buffers such as Tris or phosphate buffers to maintain a pH of about 7.0 to about 7.5. Moreover, it may be desirable to include agents which will inhibit microbial growth, such as sodium azide or Merthiolate. For extended storage in an aqueous state it will be desirable to store the solutions at 4°C, or more preferably, frozen. Of course, where the peptides are stored in a lyophilized or powdered state, they may be stored virtually indefinitely, *e.g.*, in metered aliquots that may be rehydrated with a predetermined amount of water (preferably distilled) or buffer prior to use.

5. Immunoprecipitation

The antibodies of the present invention are particularly useful for the isolation of antigens by immunoprecipitation. Immunoprecipitation involves the separation of the target antigen component from a complex mixture, and is used to discriminate or isolate minute amounts of protein. For the isolation of membrane proteins cells may typically be solubilized into detergent micelles. Nonionic salts are preferred, since other agents such as bile salts, precipitate at acid pH or in the presence of bivalent cations.

In an alternative embodiment the antibodies of the present invention are useful for the close juxtaposition of two antigens. This is particularly useful for increasing the localized concentration of antigens, *e.g.* enzyme-substrate pairs.

6. Western Blots

The compositions of the present invention will find great use in immunoblot or western blot analysis. The anti-DPD antibodies may be used as high-affinity primary reagents for the identification of proteins immobilized onto a solid support matrix, such as nitrocellulose, nylon or combinations thereof. In conjunction with immunoprecipitation, followed by gel electrophoresis, these may be used as a single step reagent for use in detecting antigens against which secondary reagents used in the detection of the antigen cause an adverse background. This is especially useful when the antigens studied are

immunoglobulins (precluding the use of immunoglobulins binding bacterial cell wall components), the antigens studied cross-react with the detecting agent, or they migrate at the same relative molecular weight as a cross-reacting signal.

Immunologically-based detection methods for use in conjunction with Western blotting include enzymatically-, radiolabel-, or fluorescently-tagged secondary antibodies against the toxin moiety are considered to be of particular use in this regard.

7. Vaccines

The present invention contemplates vaccines for use in both active and passive immunization embodiments. Immunogenic compositions, proposed to be suitable for use as a vaccine, may be prepared most readily directly from DPD peptides prepared in a manner disclosed herein. Preferably the antigenic material is extensively dialyzed to remove undesired small molecular weight molecules and/or lyophilized for more ready formulation into a desired vehicle.

The preparation of vaccines which contain peptide sequences as active ingredients is generally well understood in the art, as exemplified by U.S. Patents 4,608,251; 4,601,903; 4,599,231; 4,599,230; 4,596,792; and 4,578,770, all incorporated herein by reference. Typically, such vaccines are prepared as injectables. Either as liquid solutions or suspensions: solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified. The active immunogenic ingredient is often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, or adjuvants which enhance the effectiveness of the vaccines.

Vaccines may be conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkalene glycols or triglycerides: such suppositories may be formed from mixtures

containing the active ingredient in the range of about 0.5% to about 10%, preferably about 1 to about 2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain about 10 to about 95% of active ingredient, preferably about 25 to about 70%.

The peptides of the present invention may be formulated into the vaccine as neutral or salt forms. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the peptide) and those which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be therapeutically effective and immunogenic. The quantity to be administered depends on the subject to be treated, including, *e.g.*, the capacity of the individual's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner. However, suitable dosage ranges are of the order of several hundred micrograms active ingredient per vaccination. Suitable regimes for initial administration and booster shots are also variable, but are typified by an initial administration followed by subsequent inoculations or other administrations.

The manner of application may be varied widely. Any of the conventional methods for administration of a vaccine are applicable. These are believed to include oral application on a solid physiologically acceptable base or in a physiologically acceptable dispersion, parenterally, by injection or the like. The dosage of the vaccine will depend on the route of administration and will vary according to the size of the host.

Various methods of achieving adjuvant effect for the vaccine includes use of agents such as aluminum hydroxide or phosphate (alum), commonly used as about 0.05 to about 0.1% solution in phosphate buffered saline, admixture with synthetic polymers of sugars (Carbopol®) used as an about 0.25% solution, aggregation of the protein in the vaccine by heat treatment with temperatures ranging between about 70°C to about 101°C for a 30-second to 2-minute period, respectively. Aggregation by reactivating with pepsin treated (Fab) antibodies to albumin, mixture with bacterial cells such as *C. parvum* or endotoxins or lipopolysaccharide components of Gram-negative bacteria, emulsion in physiologically acceptable oil vehicles such as mannide mono-oleate (Aracel A) or emulsion with a 20% solution of a perfluorocarbon (Fluosol-DA®) used as a block substitute may also be employed.

In many instances, it will be desirable to have multiple administrations of the vaccine, usually not exceeding six vaccinations, more usually not exceeding four vaccinations and preferably one or more, usually at least about three vaccinations. The vaccinations will normally be at from two to twelve week intervals, more usually from three to five week intervals. Periodic boosters at intervals of 1-5 years, usually three years, will be desirable to maintain protective levels of the antibodies. The course of the immunization may be followed by assays for antibodies for the supernatant antigens. The assays may be performed by labeling with conventional labels, such as radionuclides, enzymes, fluorescents, and the like. These techniques are well known and may be found in a wide variety of patents, such as U.S. Patent Nos. 3,791,932; 4,174,384 and 3,949,064, as illustrative of these types of assays.

8. Biological Functional Equivalents

Modification and changes may be made in the structure of the peptides of the present invention and DNA segments which encode them and still obtain a functional molecule that encodes a protein or peptide with desirable characteristics. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules or receptors. Since it is the interactive capacity and nature of a protein that defines that

protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence (or, of course, its underlying DNA coding sequence) and nevertheless obtain a protein with like (agonistic) properties. Equally, the same considerations may be employed to create a protein or polypeptide with countervailing (*e.g.,* antagonistic) properties. It is thus contemplated by the inventors that various changes may be made in the sequence of DPD proteins or peptides (or underlying DNA) without appreciable loss of their biological utility or activity.

The following is a discussion based upon changing the amino acids of a protein to create an equivalent, or even an improved, second-generation molecule. The amino acid changes may be achieved by changing the codons of the DNA sequence, according to the following codon table:

TABLE 1

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		

Valine	Val	V	GUA	GUC	GUG	GUU
Tryptophan	Trp	W	UGG			
Tyrosine	Tyr	Y	UAC	UAU		

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics (Kyte and Doolittle, 1982), these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.*, still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

Two designations for amino acids are used interchangeably throughout this application, as is common practice in the art. Alanine = Ala (A); Arginine = Arg (R);

Aspartate - Asp (D); Asparagine - Asn (N); Cysteine - Cys (C); Glutamate - Glu (E); Glutamine - Gln (Q); Glycine - Gly (G); Histidine - His (H); Isoleucine - Ile (I); Leucine - Leu (L); Lysine - Lys (K); Methionine - Met (M); Phenylalanine - Phe (F); Proline - Pro (P); Serine - Ser (S); Threonine- Thr (T); Tryptophan - Trp (W); Tyrosine - Tyr (Y); Valine- Val (V).

9. Site-Specific Mutagenesis

Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent proteins or peptides, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed mutagenesis which eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double stranded vector which includes within its sequence a DNA sequence which encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared,

generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

10. DNA Segments

In other embodiments, it is contemplated that certain advantages will be gained by positioning the coding DNA segment under the control of a recombinant, or heterologous, promoter. As used herein, a recombinant or heterologous promoter is intended to refer to a promoter that is not normally associated with a DNA segment encoding a DPD peptide in its natural environment. Such promoters may include promoters normally associated with other genes, and/or promoters isolated from any viral, prokaryotic (*e.g.*, bacterial), eukaryotic (*e.g.*, fungal, yeast, plant, or animal) cell, and particularly those of mammalian cells. Naturally, it will be important to employ a promoter that effectively directs the expression of the DNA segment in the cell type, organism, or even animal, chosen for expression. The use of promoter and cell type combinations for protein expression is generally known to those of skill in the art of molecular biology, for example, see Sambrook *et al.*, 1989. The promoters employed may be constitutive, or inducible, and can be used under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins or peptides. Appropriate promoter/expression systems contemplated for use in high-level expression include, but are

not limited to, the *Pichia* expression vector system (Pharmacia LKB Biotechnology), a baculovirus system for expression in insect cells, or any suitable yeast or bacterial expression system.

In connection with expression embodiments to prepare recombinant proteins and peptides, it is contemplated that longer DNA segments will most often be used, with DNA segments encoding the entire peptide sequence being most preferred. However, it will be appreciated that the use of shorter DNA segments to direct the expression of DPD peptides or epitopic core regions, such as may be used to generate anti-DPD antibodies, also falls within the scope of the invention. DNA segments that encode DPD peptide antigens from about 10 to about 100 amino acids in length, or more preferably, from about 20 to about 80 amino acids in length, or even more preferably, from about 30 to about 70 amino acids in length are contemplated to be particularly useful.

In addition to their use in directing the expression of DPD peptides of the present invention, the nucleic acid sequences contemplated herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments. As such, it is contemplated that nucleic acid segments that comprise a sequence region that consists of at least an about 14-nucleotide long contiguous sequence that has the same sequence as, or is complementary to, an about 14-nucleotide long contiguous DNA segment of SEQ ID NO:1 or SEQ ID NO:3 will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 300, 500, 1000, (including all intermediate lengths) and even those up to and including about 4414-bp (full-length for SEQ ID NO:1) or up to an including about 4368-bp (full-length for SEQ ID NO:3), sequences, respectively, will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to DPD-encoding sequences will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are envisioned, including the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Nucleic acid molecules having sequence regions consisting of contiguous nucleotide stretches of about 14, 15-20, 30, 40, 50, or even of about 100 to about 200 nucleotides or so, identical or complementary to the DNA sequence of SEQ ID NO:1 or SEQ ID NO:3, are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 10-14 and up to about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

The use of a hybridization probe of about 14 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 14 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of about 15 to about 20 contiguous nucleotides, or even longer where desired.

Of course, fragments may also be obtained by other techniques such as, *e.g.*, by mechanical shearing or by restriction enzyme digestion. Small nucleic acid segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as PCR™, by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of DNA fragments. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively

low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating DPD-encoding DNA segments. Detection of DNA segments *via* hybridization is well-known to those of skill in the art, and the teachings of U.S. Patents 4,965,188 and 5,176,995 (each incorporated herein by reference) are exemplary of the methods of hybridization analyses. Teachings such as those found in the texts of Maloy *et al.*, 1994; Segal, 1976; Prokop, 1991; and Kuby, 1994, are particularly relevant.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or where one seeks to isolate DPD-encoding sequences from related species, functional equivalents, or the like, less stringent hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ conditions such as about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

11. Pharmaceutical Compositions

The pharmaceutical compositions disclosed herein may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard or soft shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compounds may be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 0.1%

of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 2 to about 60% of the weight of the unit. The amount of active compounds in such therapeutically useful compositions is such that a suitable dosage will be obtained.

The tablets, troches, pills, capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup of elixir may contain the active compounds sucrose as a sweetening agent methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

The active compounds may also be administered parenterally or intraperitoneally. Solutions of the active compounds as free base or pharmacologically acceptable salts can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion

medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial ad antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus ny additional desired ingredient from a previously sterile-filtered solution thereof.

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

For oral prophylaxis the polypeptide may be incorporated with excipients and used in the form of non-ingestible mouthwashes and dentifrices. A mouthwash may be prepared incorporating the active ingredient in the required amount in an appropriate

solvent, such as a sodium borate solution (Dobell's Solution). Alternatively, the active ingredient may be incorporated into an antiseptic wash containing sodium borate, glycerin and potassium bicarbonate. The active ingredient may also be dispersed in dentifrices, including: gels, pastes, powders and slurries. The active ingredient may be added in a therapeutically effective amount to a paste dentifrice that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants.

The phrase "pharmaceutically acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified.

The composition can be formulated in a neutral or salt form. Pharmaceutically acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug release capsules and the like.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, sterile aqueous media which can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage could be dissolved

in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

• • • • • • • • • •

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1

PREPARATION OF DPD ANTIGEN AND PURIFICATION OF DPD FROM HUMAN LIVER

A. MATERIALS AND METHODS

1. Materials

The materials were purchased from following sources: the TA cloning kit from Invitrogen (San Diego, CA); the pMAL protein fusion, purification expression system, and restriction enzymes were from New England Biolabs (Beverly, MA); the expression vector pBK and the random primer labeling kit from Strategene (La Jolla, CA); the coupled *in vitro* transcription and translation system and the Erase-a-base system from Promega (Madison, WI); and the 5'-AmpliFINDER RACE kit was from Clontech (Palo Alto, CA). Specific oligonucleotides were synthesized by National Biosciences (Plymouth, MN). The

hybridization membranes were obtained from Schleicher & Schuell (Keene, NH). [α-^{32}P]dCTP (3000 Ci/mmol)and [^{35}S]methionine (1000 Ci/mmol) were purchased from Amersham (Arlington Heights, IL). All other reagents were of molecular biology grade.

Polybuffer exchanger gel (PSE 94), polybuffer 74, molecular weight markers, 2', 5'-ADP-Sepharose 48, were obtained from Pharmacia (Piscataway, NJ). Coomassie brilliant blue R-250, acrylamide, and pre-stained molecular weight markers were purchased from Bio-Rad (Richmond, CA). Alkaline phosphatase labeled goat anti-rabbit antibody, nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate n-toluidine salt were obtained from Southern Biotechnology (Birmingham, AL). NADPH, FMN, and FAD were purchased from Sigma (St. Louis, MO). L-Histidine was obtained from Aldrich (Milwaukee, WI). [^{3}H]-FUra (25 Ci/mmol) was obtained from New England Nuclear Corp. (Boston, MA). [6-^{14}C]-thymine (52 mCi/mmol) were obtained from Moravek Biochemicals (Brea, CA). Radiochemicals were purified by HPLC and their purity was determined by HPLC to be $>$ 99%. All other solvents and reagents were purchased in the highest grade available.

The major buffer (buffer A) used in the preparation of this enzyme contained 35 mM potassium phosphate, 2.5 mM magnesium chloride, 10 mM 2-mercaptoethanol, pH 7.4. The equilibration buffer for the chromatofocusing column (histidine buffer) contained 25 mM L-histidine-HCl, 10 mM 2-mercaptoethanol, pH 5.7. Elution buffers for affinity column and gel filtration column were prepared from buffer A.

The design of the present study had several advantages over most previous studies of purification and characterization of this enzyme from other species. First, by introducing two new methods, chromatofocusing and HPLC gel filtration, high purity and yield of the human enzyme were obtained. Second, a specific reversed-phase HPLC method was used to determine the enzyme activity during purification and in kinetic studies. This method is a direct measure of product formation and overcomes the problems of the previous DPD assay (Porter *et al.*, 1992a; 1992b; Shiotani and Weber, 1981; Podschun *et al.*, 1989; Podschun *et al.*, 1990), which was limited both by sensitivity and specificity (Fujimoto *et al.*, 1990; Naguib *et al.*, 1985). Third, using the purified human enzyme, a polyclonal antibody was for the first time generated and shown to be specific for human liver DPD. Finally, the N-terminal amino acid sequence of the

human enzyme was determined. These new data and the availability of pure human DPD will provide a basis for further biochemical and molecular studies of the human enzyme.

2. DPD Enzyme Assay

The enzyme activity during purification was determined by measuring the catabolite of FUra using reverse-phase (HPLC (van Gennip *et al.*, 1989; Sommadossi *et al.*, 1982). The reaction mixture contained 35 mM potassium phosphate, pH 7.4, 2.5 mM magnesium chloride. 10 mM 2-mercaptoethanol, 200 μM Acro filter (Gelman Sciences, Ann Arbor, MI) and then separated by reverse-phase HPLC.

3. HPLC Analysis of Pyrimidines and Their Catabolites

Separation of pyrimidines and their catabolites wa performed by reverse-phase HPCL using a Hewlett-Packard 1050 HPLC system equipped with a filter spectrometric detector and chromatographic terminal (HP 3396 Series N Integrator). Two Hypersol® 5 mm columns (Jones Chromatography, Littleton, CO) were used in tandem as the stationery phase. Analysis of FUra and its catabodies was carried out at flow rate of 1.0 ml/min with the mobile phase containing 1.5 mM potassium phosphate, pH 8.0 with 5 mM tetrabutylammonium hydrogen sulfate. Under these conditions, typical retention times for dihydrofluorouracil and FUra were 9 and 21 min. respectively.

Using the same stationary phase as above, analysis of thymine and catabolites was carried out at a flow rate of 0.5 ml/min with the mobile phase containing 1.5 mM potassium phosphate, pH 8.4, with 5 mM tetrabutylammonium hydrogen sulfate. Under these conditions, typical retention times for dihydrothymine and thymine were 22 and 27 min. respectively. Analysis of uracil and its catabodies was also carried out using the same HPLC system fro analysis of thymine and its catabolites, with typical retention times of 13 and 19 min for dihydrouracil and uracil, respectively.

4. SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE was carried out in a 1.0 mm thick, 7% (w/v) polyacrylamide gel containing 0.375 M Tris-HCl (pH 8.8) and 0.1% SDS. Samples were prepared by mixing them with an equal volume of sample buffer (0.0625 M Tris-HCl, pH 6.8; 10% glycerol; 0.2% SDS (w/v); 80 mM 2-mercaptoethanol) and boiling for 5 minutes. Electrophoresis was conducted at a constant current of 30 mA for 30 min at 25°C.

Gradient SDS-PAGE was carried out in a 1.0 mm thick, 4-20% gradient gel (Bio-Rad Mini-Protean II). Samples were prepared by mixing them with four volumes of the above sample buffer and boiling for 5 min. The electrophoresis was conducted following the manufacture's instruction, at a constant voltage of 200 V for 60 min at 25°C.

5. Native Polyacrylamide Gel Electrophoresis

Native gel electrophoresis was carried out in a 1.0 mm thick, 9% (w/v) polyacrylamide gel containing 0.06 M Tris-HCl (pH 8.8), with 0.0025 % (w/v) riboflavin phosphate. Sample were prepared by mixing them with an equal volume of sample buffer (40% sucrose, 10 mM 2-mercaptoethanol). The electrophoresis was conducted at a constant current of 30 mA for 30 min at 4°C.

6. Staining Procedure

a. Coomassie blue: The gel was fixed in a 5% methanol / 7% acetic acid solution for 30 min and stained overnight using 0.01% (w/v) Coomassie brilliant blue R-250 in a 5% trichloroacetic acid / 2.5% methanol / 3.5% acetic acid solution.

b. Silver Staining: The gel was fixed in 40% methanol / 10% acetic acid for 40 min and then stained using the Bio-Rad silver stain (Merril *et al.*, 1981). Briefly, following fixation, the gel was incubated in oxidizer solution for 20 min. The gel was then washed with distilled deionized water and incubated with silver solution for 30 min. The gel was again washed with distilled deionized water and incubated with the developing solution supplied by the manufacturer.

7. Electroelution From Native Polyacrylamide Gel

Gel electrophoresis was carried out on 200 μg purified DPD under non-denaturing conditions in a 9% (w/v) polyacrylamide gel. This strip was lined up with the unstained gel and the single corresponding band cut out of the unstained gel. The gel was minced and electroeluted in a Bio-Rad Model 422 electro-eluter in 25 mM Tris/192 mM glycine buffer, pH 8.3, contain 5% glycerol, 5 mM 2-mercaptoethanol for four hours at 10 mA (constant current) at 4°C. The sample was then dialyzed overnight at 4°C in 1 liter of buffer A, pH 7.4, before being assayed. Other fractions from the gel were treated in the same way.

8. Molecular Weight Determination

The molecular weight of native DPD was determined by HPLC gel filtration. A 2.15 x 60 cm TSK-250 gel filtration HPLC column (Bio-Rad) was equilibrated with buffer A, pH 7.4, at a flow rate of 2.5 ml/min. The column was calibrated using known molecular weight standards and the retention time of individual proteins determined by their peaks of absorbance at 280 nm. The retention time of purified enzyme was then compared to those of the molecular weight standards. The molecular weight of reduced, denatured DPD was determined by SDS-polyacrylamide gel electrophoresis, using standard proteins of known molecular weights.

9. Flavin Determination

The purified enzyme dissolved in 35 mM potassium phosphate pH 7.4, 2.5 mM $MgCl_2$and 5 mM 2-mercaptoethanol was boiled in a water bath for 10 min in the dark to release flavin. After removing the precipitate by centrifugation, aliquots of supernatant were analyzed qualitatively for flavin composition by HPLC separation on a reverse phase c_{16} column with a linear gradient (0-66% methanol) in 20 mM potassium phosphate, pH 5.6, at a flow rate of 1 ml/min at 25°C. Flavins were detected by their absorbance at 230 nm. The FAD/FMN composition of the supernatant was analyzed quantitative by fluorescence measurements at different pH values (Faeder and Siegel, 1973) with FAD and FMN standards purified on DEAE-cellulose (Massey and Swoboda, 1963).

10. Metal and Sulfide Determination

The metal content of purified DPD was determined by atomic absorption spectrophotometry. Acid-labile sulfide was measured by the methylene blue method (Rabinowitz, 1978).

11. Kinetic Studies

Initial reaction rates were determined at various concentrations of each substrate (0.5, 1, 2, 3, 4, 5, 7.5, 10, 20, 40, 80, 100, 200, 500, 1000 μM) in the presence of 200 μM NADPH. Kinetic studies for NADPH were carried out at various concentrations of NADPH (0.5, 1, 2, 3, 5, 7.5, 10, 20, 40, 60, 80, 100, 200, 500, 1000 μM) in the presence of 20 μM uracil, thymine or FUra. Reactions were run in buffer A at 37°C. The incubation time and protein concentration were adjusted so that no more than 10%

of the limiting substrate was consumed. Estimation of the apparent K_m and V_{max} values for each substrate was performed by fitting these data for several concentrations of different substrates to the Michaelis-Menton equation by non-linear regression analysis (Cleland, 1979). Enzyme activity is expressed as μmol of product formed per min per mg of protein.

12. Amino Acid Analysis

Protein samples were carboxymethylated as described (Allen, 1989). The amino acid composition was determined by first hydrolyzing the sample at 100°C for 20 hours in the presence of 6 N HCl: the amino acids were analyzed by reversed-phase HPLC using the PICO TAG system (Waters Associates, Milford, MA).

13. *N*-Terminal Amino Sequencing

Determination of amino-terminal sequences of purified human liver DPD was performed by automated Edman degradation with a gas-phase sequencer (Model 470A, Applied Biosystems, Forster City, CA). Phenylthiohydantoin derivatives of amino acids were separated by HPLC using an RP 18 column (Matsudaira, 1987).

14. Preparation of Polyclonal Antibody

Male New Zealand rabbits were immunized with subcutaneous injections of purified DPD. The first injection consisted of 50 μg of purified antigen mixed with an equal volume of Freund's complete adjuvant. Two weeks later, these rabbits were injected with the antigen (50 μg) mixed in an equal volume of Freund's incomplete adjuvant; three weeks following the second injection, this injection was repeated. Aliquots of serum samples from ear nicks were screened for antibody formation using enzyme-linked immunosorbent assay (Gaastra, 1984) and Western blot analysis (Towbin *et al.*, 1979). Two weeks following the third injection, the rabbits were sacrificed by cardiac puncture, and their blood collected. To allow the blood to clot, the sample was incubated at 37°C for 60 min, left at room temperature for 4 h and then kept at 4°C overnight. The clot was gently removed, and the serum was centrifuged at 2000 RPM for 15 min. The serum was loaded on a 1 × 10 cm protein A-Sepharose 4 Fast Flow column (Sigma Chemical Co., St. Louis, MO), previously equilibrated with phosphate-buffered saline. The column was washed with 4 column volumes of phosphate-buffered

saline, and the IgG antibodies were eluted with an acid wash consisting of 0.2 M glycine-HCl containing 0.075 M NaCl, pH 2.5. Immediately upon elution from the column the fractions were neutralized with 1.0 M Tris-HCl, pH 10.

15. Immunoblot Analysis

SDS-PAGE on a 4-20% gradient gel was performed using freshly prepared 100,000 x *g* human liver supernatant and purified human liver DPD. The proteins were transferred from the gel to a nitrocellulose filter (Towbin *et al.*, 1979). The nitrocellulose filter was incubated overnight at 4°C with the polyclonal antibody (Ig G) purified by protein A column (diluted 1:2000) in a 120 mM borate-saline solution containing 1% (w/v) BSA, pH 8.5. The nitrocellulose filter was washed with borate-saline containing 0.1% Tween-20™ (w/v) and incubated with a secondary, alkaline phosphatase-labeled goat anti-rabbit antibody. The location of immunoreactive proteins on the nitrocellulose filter was developed in a 0.1 M sodium carbonate buffer (100 ml, pH 9.5) containing 30 mg nitro blue tetrazolium (added as a 1 ml solution dissolved in 70% dimethylformamide) and 15 mg 5-bromo-4-chloro-3-indolyl phosphate *p*-toluidine salt (added as a 1 ml solution dissolved in 100% dimethylformamide).

16. Protein Determination

The amount of protein in the sample was determined using Bio-Rad protein determination reagent with BSA as a standard (Lowry *et al.*, 1951).

17. Enzyme Purification

All procedures were performed at 4°C. The summary of purification is listed in Table 2.

Fraction 1 (Preparation of Crude Extract): Human liver (received from the National Disease Research Interchange through an Institutionally approved protocol) was removed from transplant donors as soon as possible after cessation of cardiac function. The tissue was cut into 250 gram pieces, perfused with cold saline, and frozen at 70°C. Twenty-four hours prior to use, liver was placed in a paper-lined ice bucket and set in a 4°C cold room. The partially thawed liver was minced and homogenized in four volumes of buffer A, in the presence of 0.25 M sucrose, 1 mM benzamidine, 1 mM

aminoehylisothiouronium bromide, and 5 mM EDTA. The homogenate was centrifuged at 100,000 × *g* for 60 min in order to obtain a cytosolic fraction.

Fraction 2 (Acid Precipitation): Acetic acid was added to the resulting supernatant of homogenate (Fraction 1) to adjust pH to 4.85 with constant stirring for 15 min at this pH. The enzyme solution was then centrifuged at 30,000 × g for 30 min. The supernatant was removed and adjusted with 0.5 N KOH to pH 7.4.

Fraction 3 (Ammonium Sulfate Fractionation): Solid ammonium sulfate was slowly added to fraction 2 until a 33% saturation was obtained. The mixture was stirred for 30 min and then centrifuged at 30,000 × *g* for 30 min. Additional ammonium sulfate was added to the supernatant until a 55% saturation was obtained with constant stirring for 30 min. The enzyme solution was then centrifuged at 30,000 × *g* for 30 min. The precipitate was dissolved in 25 mM histidine-HCl buffer, pH 5.7, and dialyzed overnight against 10 liter of the same buffer.

Fraction 4 (Chromatofocusing): The dialyzed sample from Fraction 3 was centrifuged at 30,000 × *g* for 30 min and then loaded onto a chromatofocusing column (1.6 × 100 cm) packed with PBE-94 previously equilibrated with 25 mM histidine-HCl buffer, pH 5.7. The column was re-equilibrated with 5 column volumes of the equilibration buffer. The column was then eluted by a polybuffer 74 diluted 1:8 with distilled deionized water (final pH adjusted to 4.0 with HCl), in the presence of 10 mM 2-mercaptoethanol.

Fraction 5 (2',5' ADP-Sepharose 4B Affinity Chromatography): The polled fractions with DPD activity from the chromatofocusing column were concentrated by Amicron centriprep 10 concentrator and loaded onto a 2', 5'-ADP-Sepharose 4B affinity column (1 × 40 cm) previously equilibrated with buffer A. The column was washed with 20 column volumes of buffer A. 10 column volumes of 50 mM KCl-buffer A, 10 column volumes 100 mM KCl-buffer A, 2 column volumes of 200 mM KCl-buffer A. Enzyme activity was eluted with 0.1 mM NADPH in buffer A. Fractions containing DPD activity were pooled and concentrated in an Amicron Centricon 10 concentrator.

Fraction 6 (Gel Filtration Chromatography): The pooled, concentrated fractions with DPD activity from the affinity column were injected onto a Biorad TSK-250 gel

filtration column (2.15 × 60 cm), previously equilibrated with buffer A. Enzyme activity was eluted by buffer A in a flow rate or 2.5 ml/min. Fractions containing DPD activity were pooled and concentrated in an Amicron centricon 10 concentrator.

B. RESULTS

1. Enzyme Purification

In the present study, DPD activity was purified from the soluble fraction of homogenized frozen human liver. Initially, the 100,000 × *g* human liver supernatant fraction was precipitated by addition of acetic acid to pH 4.85 followed by ammonium sulfate fractionation. After 55% ammonium sulfate precipitation, the pellet was resuspended and dialyzed against 25 mM histidine-HCl buffer, pH 5.7, overnight and then loaded onto a PBE-94 chromatofocusing column equilibrated with the same buffer. The column was eluted by polybuffer 74, creating a pH gradient from pH 5.6 to 4.0. DPD activity was subsequently eluted at pH 4.6 (± 0.2). Fractions containing DPD activity were pooled, concentrated, and loaded onto a 2', 5'-ADP-Sepharose 4B affinity column; proteins which did not bind and those loosely bound to the affinity matrix were sequentially eluted with buffer A and an increasing step gradient of buffer A containing 50 mM, 100 mM, and 200 mM KCl, respectively. DPD activity was recovered from the affinity column by elution with 0.1 mM NADPH. Concentrated, affinity-purified DPD activity was then chromatographed on an HPLC gel filtration column which separated DPD activity from other protein contaminants. In a typical preparation, the final product had a 7800-fold enrichment of enzyme activity, with an overall recovery of 20% (Table 2).

TABLE 2

PURIFICATION OF DPD FROM HUMAN LIVER

Step	Total Protein	Total activity[a]	Specific activity[a]	Recovery[a]	
	mg	*nmol/min*	*nmol/min/mg*	*%*	*-fold*
Crude supernatant	24229	4911	0.2027		
pH 4.85 treatment	15770	4173	0.2646	85.0	1.3
Ammonium sulfate fractionation[b]	5719	4158	0.7271	84.7	3.6
Chromatofocusing	193	3271	16.981	66.6	83.8
2',5'-ADP-Sepharose affinity	1.97	1447	734.50	29.5	3624
Gel filtration	0.63	999	1585.9	20.3	7824

[a] All values calculated using FUra as a substrate.

[b] After dialysis and centrifugation.

2. Molecular Weight Determination

Purified enzyme was homogeneous as judged by HPLC gel filtration on a TSK 250 column (calibrated with known standards) showing a single, symmetrical peak corresponding to a molecular mass of 210±5 kDa, which was not influenced by the presence of 2-mercaptoethanol. The homogeneity of purified human liver DPD was also determined by native gel electrophoresis. Under nondenaturing conditions, a single band was obtained from the native gel by staining with either Coomassie Blue R-250 or a silver-staining technique. Following electroelution from the gel, DPD activity was recovered from the single band. No enzyme activity was detected from other fractions of the gel. The denatured, reduced enzyme gave two sharp protein bands with apparent molecular masses of 105 and 90±3 kDa on a 7% SDS-polyacrylamide gel. To further characterize this enzyme under denaturing conditions, purified human liver DPD was examined using a 4-20% gradient SDS gel. With silver staining, three different bands with molecular masses of 105, 90, and 15 kDa were observed. The binding capacity of

the 15-kDa band for Coomassie Blue R-250 was very low, but this band was readily detected by silver staining.

3. Determination of the Isoelectric Point of DPD

Elution from the chromatofocusing column demonstrated an apparent isoelectric point (pI) of 4.6 (± 0.2) (Table 3). The elution pattern was symmetrical, further suggesting that the purified human liver DPD was homogeneous.

TABLE 3

COMPARISON OF HEPATIC DPD FROM HUMAN, PIG, AND RAT

Parameter	Liver from		
	Human	Pig (28)	Rat (24, 25)
Molecular mass (kDa)	210	206	220 (207)[a]
pI	4.60	4.65	5.25
Iron (mol/mol enzyme)	33.2	30.4	3.0 (14.0)
Inorganic sulfur (mol/mol enzyme)	31.6	31.3	NA[b]
FMN (mol/mol enzyme)	1.50	1.7	NA (0.7)
FAD (mol/mol enzyme)	1.51	1.6, 1.9	3.75 (0.76)

[a] Values in parentheses from Ref. 25.

[b] NA, data not available.

4. Flavin Determination

Purified human liver DPD had an amber color (in buffer A) and showed the characteristic absorption spectrum of a reduced flavoprotein. The nature of the flavin cofactor in the enzyme molecule was shown by HPLC to be FAD and FMN. No conversion of FAD to FMN was detectable under these conditions. FAD and FMN were quantitated by a simultaneous fluorometric assay. As illustrated in Table 3, human DPD contains approximately 2 mol each of FAD ad FMN per mol of enzyme.

5. Metal and Sulfide Determination

To determine the metal content in this enzyme, purified human liver DPD was subjected to atomic absorption spectrometry. Approximately 33 mol of iron per mol of enzyme were detected; no iron was detectable in the buffers used in the purification procedure. No other metal ions were found in purified enzyme. The acid-labile sulfide content of purified human liver DPD was analyzed to determine the binding mode of the iron atoms. As shown in Table 3, the acid-labile sulfide content was almost equal to the iron content suggesting the presence of Fe-S centers in purified DPD; no sulfide was detected in the buffers used in this study.

6. Amino Acid Composition

The amino acid composition of carboxymethylated DPD was determined and the results are presented in Table 4. These data represent the mean of four separate DPD preparations. The amino acid compositions of rat and pig liver DPDs were identical to the human sequence.

TABLE 4

AMINO ACID COMPOSITIONS OF HEPATIC DPD FROM HUMAN, PIG, AND RAT

Amino acid	Amount (residues/mol enzyme) from liver of		
	Human	Pig (28)	Rat (24)
Asp + Asn	163.5	177.7	185.1
Glu + Gln	189.8	196.4	141.2
Ser	108.7	119.4	121.9
Gly	180.5	185.4	121.1
His	38.7	22.0	22.3
Arg	92.6	72.7	63.0
Thr	130.3	109.6	102.7
Ala	185.1	171.5	121.9
Pro	147.3	126.2	113.3
Tyr	66.0	30.3	35.9
Val	140.4	108.4	96.2
Met	51.9	94.0	41.8
Cys	16.2	30.7	27.9
Ile	115.7	106.9	99.9
Leu	199.7	167.4	134.1
Phe	92.2	74.3	64.1
Lys	94.3	124.1	92.0
Trp	*ND*	*ND*	*ND*

ND = Not Determined.

7. *N*-terminal Amino Sequence

The N-terminal amino residues of the 105- and 90-kDa peptides, following separation on a 7% SDS-PAGE, were identical and the same as that of native enzyme (Table 5).

TABLE 5

AMINO-TERMINAL AMINO ACID SEQUENCES OF DPD FROM HUMAN LIVER

Sample	Residue									
	1	2	3	4	5	6	7	8	9	10
Native enzyme[a]	Val	Leu	Ser	Lys	Asp	Ser	Ala	Asp	Ile	Glu
105-kDa band[b]	Val	Leu	Ser	Lys	Asp	Ser	Ala			
90-kDa band[b]	Val	Leu	Ser	Lys	Asp	Ser	Ala			

[a] After HPLC gel filtration.

[b] After separation on SDS-PAGE (7%).

8. Optimization of pH ad Temperature Conditions

In a series of 11 mM potassium phosphate buffers covering a pH range between 4.0 and 9.0 with FUra as a substrate, the highest DPD activity was observed at pH 7.4. Similarly, when incubated at temperatures over a range between 4.0 and 70.0°C, the highest DPD activity was observed at 37°C.

B. RESULTS

1. Kinetic Properties

Table 6 summarizes the kinetic studies of purified human liver DPD, with comparison to rat and pig liver enzymes. Using standard assay conditions at pH 7.4 and 37°C, in the presence of 200 μM NADPH, enzyme kinetic studies revealed apparent K_m values for uracil, thymine, and FUra of 4.9, 4.8 and 3.3 μM, with corresponding V_{max} values of 0.6, 0.7, and 0.9 μmol/min/mg protein, respectively. Under the above

conditions, substrate inhibition was observed for all substrates examined in the study. In the presence of 20 μM pyrimidine substrate, apparent K_m values for NADPH were 9.6 μM with uracil, 15.8 μM with thymine, and 10.1 μM with FUra, respectively. Under these conditions, no significant inhibition by NADPH was observed.

2. Immunological Characterization

In the present study, rabbit polyclonal antibody was generated against purified human liver DPD. Using this antiserum immunoblot analysis of proteins in 100,000 × g human liver supernatant, after separation on SDS-PAGE (4-20% gradient), revealed a single 105-dKa band. Preimmune serum from the same rabbit did not detect any band under the same conditions.

TABLE 6

COMPARISON OF KINETICS FOR HEPATIC DPD FROM HUMAN, PIG AND RAT

Substrate	Parameter	Liver from		
		Human	Pig (28)	Rat (24)
Uracil	K_m (μM)	4.9	1.98	1.80
	V_{max} (μmol/min/mg)	0.6	0.33	0.69
Thymine	K_m (μM)	4.8	2.66	2.6
	V_{max} (μmol/min/mg)	0.7	0.25	0.49
FUra	K_m (μM)	3.3	5.50	NA*
	V_{max} (μmol/min/mg)	0.9	0.4	NA
NADPH	K_m (μM) with uracil	9.6	11.36	11
	K_m (μM) with thymine	15.8	NA	15
	K_m (μM) with FUra	10.1	NA	NA

* NA, data not available.

Compared to previous studies on purification of this enzyme from other species, the present study utilized a novel procedure and represents a 5-fold improvement on previous methods of purification of this enzyme from rat liver (Shiotani and Weber, 1981)

and 2-fold from pig liver (Podschun *et al.*, 1989). A specific polyclonal antibody has been raised for the first time against human liver DPD.

Purification of human liver DPD to homogeneity was accomplished by a combination of acid precipitation, ammonium sulfate fractionation, chromatofocusing, affinity chromatography, and HPLC gel filtration. The final product, when analyzed by native PAGE, consisted of a single protein band. Following electroelution from the nondenaturing gel, the band was shown to have DPD activity. The degree of homogeneity of the native enzyme was demonstrated by the symmetry of the single peak (absorbance and DPD activity) by HPLC gel filtration. Further confirmation of the homogeneous nature of purified human liver DPD was obtained using a polyclonal antibody raised in rabbits against purified enzyme.

When purified DPD was resolved by SDS-PAGE on a 4-20% gradient gel, three polypeptide bands, with molecular masses of 105, 90, and 15 kDa, were observed.

The following data suggest that native human liver DPD consist of two 105-kDa subunits with the 90- and 15-kDa polypeptides representing degradation products: first, under nondenaturing conditions purified enzyme eluted during HPLC gel filtration as one symmetrical peak which corresponded to DPD activity; second, DPD activity was recovered following electroelution from the single band of the native gel (no other proteins and no enzyme activity were detected from other fractions of the native gel); third, fractions from chromatofocusing, affinity, and HPLC gel filtration columns which had DPD activity were shown on SDS-PAGE to contain the 105-, 90-, and 15-kDa polypeptides (other fractions without DPD activity did not contain any one of these three polypeptides); fourth, N-terminal amino residues from native DPD (210 kDa) and from 105- and 90-kDa polypeptides were identical; and fifth, immunoblot analysis using the rabbit polyclonal antibody detected a single 105-kDa protein band with the crude human liver cytosol, whereas three bands with molecular masses of 105, 90, and kDa were detected with purified DPD.

Determination of the isoelectric point (pI) of purified enzyme revealed a lower pI for human liver DPD compared with rat liver DPD (Shiotani and Weber, 1981) (pI 5.25).

In this respect, human liver DPD is more like pig liver enzyme (Podschun *et al.*, 1989) with a similar pI (4.60 *vs.* 4.65).

Comparison of the amino acid composition of DPD from three mammalian species (human, rat, and pig) demonstrated that a significant deviation in composition occurred for acidic amino acids (more abundant in human liver DPD). Human DPD has approximately twice as many histidine residues as rat and pig DPD.

The amber color (in buffer A) and the characteristic absorption spectrum of human liver DPD suggest it is a flavoprotein. Equal amounts of FMN and FAD were detected in purified enzyme. Similar results were reported for pig liver DPD (Podschun *et al.*, 1989). In contrast, only FAD was found in rat liver DPD (4 mol per mol of enzyme) (Shiotani and Weber, 1981). Both FAD and FMN have been reported in rat liver DPD (1 mol of each flavin per mol of enzyme) (Fujimoto *et al.*, 1990). The role of flavins in this enzyme is unclear. It has been suggested that flavin may regulate the enzyme half-life or synthesis.

Determination of metal and acid-labile sulfide contents of human liver DPD revealed similar amounts per mol of enzyme, suggesting the presence of Fe-S centers. Purified human liver DPD contained 8 mol each of iron and acid-labile sulfide per mol flavin nucleotide. These results were in agreement with the report on pig liver enzyme (Podschun *et al.*, 1990). However, the results of iron determination for rat liver enzyme from different preparations varied: one report suggested only 3 mol of iron per mol of enzyme was present (Shiotani and Weber, 1981), while another reported 14 mol of iron per mol of enzyme (Fujimoto *et al.*, 1990).

In most of the previous DPD purifications from other species (Shiotani and Weber, 1981; Podschun *et al.*, 1989; Podschun *et al.*, 1990), enzyme activity was determined by the decrease in NADPH assessed by measuring changes in absorbance at 340 nm. This method is limited in both sensitivity and specificity, particularly in the first several steps of purification where more than one enzyme consumes NADPH.

The enzyme activity in the present study was quantitated by measuring specific product formation. Using HPLC methodology, kinetic studies have demonstrated similar kinetic properties for the natural substrates, uracil and thymine. Significant substrate

inhibition was observed for uracil, thymine, and FUra at 100 μM or higher. Substrate inhibition was reported with purified pig liver DPD (Podschun *et al.*, 1989; Podschun *et al.*, 1990) and crude extracts of some human tissues (Naguib *et al.*, 1985). However, no substrate inhibition was reported with rat liver DPD (Shiotani and Weber, 1981). In the presence of 20 μM of each pyrimidine substrate, saturation of enzyme activity was detected at 30 μM NADPH, but significant inhibition by NADPH was not observed until 1000 μM NADPH. In the present study, FUra was the preferred substrate for human liver DPD compared to uracil and thymine. It is possible that the variations in estimated kinetic parameters for different species may result from several factors, including species differences, varying methods in determination of enzyme activity, and varying degrees of purification.

EXAMPLE 2

CLONING, CHARACTERIZATION, AND EXPRESSION OF BOVINE LIVER DPD

A. MATERIALS AND METHODS

1. Materials

The materials were purchased from following sources: the TA cloning kit from Invitrogen (San Diego, CA); the pMAL protein fusion, purification expression system, and restriction enzymes were from New England Biolabs (Beverly, MA); the expression vector pBK and the random primer labeling kit from Stratagene (La Jolla, CA); the coupled *in vitro* transcription-translation system and the Erase-a-base system® from Promega (Madison, WI); and the 5'-AmpliFINDER RACE kit was from Clontech (Palo Alto, CA). Specific oligonucleotides were synthesized by National Biosciences (Plymouth, MN). The hybridization membranes were obtained from Schleicher & Schuell (Keene, NH). [α-^{32}P]dCTP (3000 Ci/mmol)and [^{35}S]methionine (1000 Ci/mmol) were purchased from Amersham (Arlington Heights, IL). All other reagents were of molecular biology grade.

2. Partial Amino Acid Sequence of Purified Bovine Liver DPD

Bovine liver DPD was purified as previously described (Lu *et al.*, 1993). N-terminal amino acid sequence was obtained directly from the purified enzyme as described earlier (Lu *et al.*, 1992). Internal amino acid sequence was obtained from peptides

generated by digestion of DPD with cyanogen bromide (CNBr) as follows. Approximately 100 μg of purified bovine liver DPD was placed into a clean 1.5 ml microcentrifuge tube containing 100 μl of 10 mg/ml fresh cyanogen bromide (Kodak, Rochester, NY) in 70% formic acid. The reaction was incubated for 12 hours at room temperature in the dark. The mixture was then diluted with ten volumes of distilled water, frozen at -70°C, and lyophilized. The cyanogen bromide-generated polypeptides were resolved by SDS-PAGE and electrophoretically transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore, Inc., Bedford, MA) (Matsudaira, 1987). The membrane was rinsed in distilled water and the protein visualized by Ponceau-S staining. A single, well-separated band corresponding to a molecular weight of approximately 30 kDa was excised for analysis. The sequence of the purified, CNBr-generated fragment was determined by the Protein Analysis and Peptide Synthesis Core Facility in the Comprehensive Cancer Center at University of Alabama at Birmingham using an Applied Biosystems Model 470 protein sequencer with an on-line 120A PTH analyzer. Analysis of two separate samples generated identical sequence data.

3. cDNA Synthesis

Bovine liver was obtained directly from a local slaughter house and snap frozen in dry ice/methanol. Total RNA was isolated by the method of Ausebel (Ausebel *et al.*, 1987). Purification of poly(A)$^+$ RNA was performed using an Oligotex-dT mRNA kit (Qiagen) according to the manufacturers instructions. cDNA synthesis was performed in a 20 μl reaction volume containing 200 units of Moloney murine leukemia virus reverse transcriptase (Promega), the enzyme buffer (as supplied by the manufacturer), 1 μg of poly(A)$^+$ RNA, 20 units of RNasin (Promega), dNTPs (1 mM each), and 0.5 μg of one of the following primers: oligo(dT) (Promega); specific primers generated from bovine DPD cDNA; or random hexamers. Following a one hour incubation at 37°C, the reaction mix was diluted to a final volume of 1.0 ml with ddH_2O and stored at -70°C in 100 μl aliquots.

4. Amplification and Subcloning of the PCR™ Products

The amplification of DPD cDNA was performed in a 50 μl reaction volume containing 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 2.5 mM $MgCl_2$, plus dNTPs (0.2 mM

each), 50 pmol of each primer and 5 μl template cDNA (see cDNA synthesis above), 2.5 U of *Thermus aquaticus* (Taq) polymerase (Perkin-Elmer/Cetus), and overlaid with 50 μl of mineral oil. The samples were amplified in a PTC-100 thermal cycler (MJ Research, Inc., Watertown, MA) programmed for a temperature-step cycle of 94°C (1 min), 52°C (2 min), and 72°C (3 min). This cycle was repeated for a total of 40 cycles with a 10 min extension at 72°C after the final cycle. The reaction products were purified from the rest of the reaction mixture by electrophoresis in low-melting-point agarose (USB, Cleveland, OH). Bands were visualized with an ultraviolet transilluminator after staining with ethidium bromide, excised from the gel, and placed at 65°C for 5 min. The samples were then purified by phenol/choloform extraction followed by ethanol precipitation. The purified PCR™ products were subcloned directly into the pCRII® vector (Invitrogen) following the instructions supplied by the manufacturer.

5. Cloning Strategy and Amplification of Bovine Liver DPD cDNA

A mixed oligonucleotide-primed amplification of cDNA (MOPAC) strategy was utilized to amplify bovine liver DPD cDNA by PCR™ (Lee *et al.*, 1988). The amplification of the full-length DPD cDNA was accomplished in four stages:

1) The PCR reaction mixture used oligo(dT) primed bovine liver cDNA as the template. Two degenerate oligonucleotide primers were designed based on the sequence of a 23 amino acid tryptic fragment (KAEASGAXALELNLSCPHGMGER) (SEQ ID NO:7) generated from purified bovine liver DPD (Porter *et al.*, 1992a; 1992b). Primer A (sense: 5'-AARGGIGARGCITCIGGIGC-3') (SEQ ID NO:8) and primer B (antisense: 5'-TCICCCATICCRTGIGG-3') (SEQ ID NO:9) corresponding to amino acid sequences KAEASGA (SEQ ID NO:10) and PHGMGE (SEQ ID NO:11) respectively, were used to amplify a 65 base pair product (FIG. 1A). Amplified products were resolved on a 2% low melt agarose gel and subcloned into pCRII®.

2) To obtain the 3' end of the clone, first-strand cDNA synthesis was carried out using an adaptor-oligo $(dT)_{17}$ primer (5'-ACTCGATGCGACATCGATTTTTTTTTTTTTTTT-3') (SEQ ID NO:12) containing a *Sal* site. Amplification was performed using primer C (sense: 5'-CGCCCTGGAGTTAAATTTATCGTG-') (SEQ ID NO:13) designed to the region coding for

the known amino acid sequence, DALELNLSC (SEQ ID NO:5) and from the specific sequence data generated from the amplified region between primers A and B), together with the adaptor region of the oligo (dT) primer (FIG. 1B). The amplification was performed as described previously (Frohman *et al.*, 1988) with an annealing temperature of 56°C.

3) To obtain the 5' end of the clone, the first 16 amino acids (VLSKDVADIESILALN) (SEQ ID NO:14) from the amino-terminal end of purified bovine liver DPD were determined. A degenerate oligonucleotide, primer D (sense: 5'-AARGAYGTIGCIGATATCGA-3') (SEQ ID NO:15), was designed to the portion of the N-terminal amino acid sequence KDVADIE (SEQ ID NO:6). Sequence data obtained from the 2360 base pair fragment (FIG. 1B) was used to design primer E (antisense: 5'-AACCCAGCGACAGATGTTCC-3') (SEQ ID NO:16) and amplification carried out with an annealing temperature of 47°C (FIG. 1C). The resulting PCR™ product (2076 base pairs) was purified and subcloned into pCRII® as previously described.

4) To extend the nucleotide sequence toward the initiating ATG, rapid amplification of the cDNA 5' end (RACE) method was applied (Frohman, 1990) using a 5'-AmpliFINDER RACE kit following the manufacturer's instructions. The specific antisense primer (5'-GTCGTGTGCTTGATGTCATC-3') (SEQ ID NO:17) was used for first-strand cDNA synthesis followed by PCR™ amplification with the specific antisense primer (5'-GCTTCTCGCAATTAAAGCAG-3') (SEQ ID NO:18). The sense primer (5'-CCTCTGAAGGTTCCAGAATCGATAG-3') (SEQ ID NO:19) was complementary to the anchor sequence utilized in the 5'-AmpliFINDER RACE kit (FIG. 1D). The resulting 237 base pair PCR™ product was subcloned and sequenced.

To facilitate expression studies, the three RT-PCR™ fragments (237, 2076, 2360 base pairs) were ligated together to form a complete 4414 base pair DPD clone. The 2360 base pair fragment was ligated to the 2076 base pair fragment at an overlapping *Bam*HI site to form a new 4330 base pair construct. To obtain the complete full-length clone, the 237 base pair 5'-end was ligated to the 4330 base pair construct at an overlapping *Msc*I site. Using standard DNA recombination methods (Sambrook *et al.*, 1989), the complete 4414 base pair cDNA was subcloned into the pCRII® plasmid.

6. Northern Blot Analysis of Rat Liver RNA

Total and poly(A)$^+$ RNA were prepared from bovine liver by the methods described above for cDNA synthesis. Radiolabeled probe (specific activity = 1 × 10^{11} cpm/μg) was prepared with a Pharmacia Oligolabelling Kit using full-length 4414 base pair bovine liver cDNA as the template. Total RNA (30 μg) and poly(A)$^+$ RNA (1 μg) were resolved by electrophoresis in a 1.5% agarose-formaldehyde denaturing gel and transferred to a NYTRAN™ nylon membrane (Schleicher & Schuell). The filters were UV cross-linked, prehybridized for 30 min and then hybridized for two hours at 60°C in 10 ml QuickHyb™ solution (Stratagene). The filters were washed under stringent conditions following the manufacturer's instructions.

7. *In vitro* Transcription-Translation

The full-length 4414-bp bovine liver cDNA was constructed in the pCRII® vector downstream from the SP6 RNA polymerase promoter. *In vitro* transcription and translation was conducted with the TNT™ SP6 coupled reticulocyte lysate system (Promega) using [*35*S]methionine for labeling of the synthesized proteins. The translated products were resolved by SDS-PAGE in an 8% polyacrylamide gel (Lu *et al.*, 1993). The gels were vacuum-dried at 65°C and exposed to autoradiography film for 6 hr.

8. Generation of Bovine DPD Prokaryotic Expression Vector

For the expression of bovine liver DPD in *E. coli*, the 4330 base pair bovine liver DPD cDNA (not including the initiating ATG) was subcloned into the *Eco*RI-*Sal*I sites of the bacterial expression vector pMal-c2. The pMal-c2 vector will express the heterologous cDNA as a maltose binding protein fusion protein. The DPD cDNA was PCR™ amplified while in the pCRII® vector using the mutated sense primer (5'-CTGGAATTCGGCTT**A**AAGGACGTGGCGG-3') (SEQ ID NO:20) along with the adaptor region of the oligo (dT) as the antisense primer. To generate the appropriate reading frame, an additional adenosine base (bold face and underlined) was incorporated in the mutated primer just before the cDNA coding sequence. The PCR™ product was digested with *Eco*RI-*Sal*I and directionally subcloned into corresponding sites in the pMAL-c2 expression vector.

9. Bacterial Expression of Bovine Liver DPD

The pMal-c2/bovine DPD construct was transfected into *E. coli* KS 1000 cells made competent using a $CaCl_2$ procedure (Davis *et al.*, 1986). Bacterial colonies containing plasmids were selected by ampicillin resistance and the presence of the mutated bovine cDNA confirmed by restriction mapping and sequence analysis. For expression and induction of DPD, transformed KS 1000 cells were grown at 37°C in Luria broth containing 50 μg/ml ampicillin. When the culture reached an A_{600} of 0.5; isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 0.3 mM. Following a further two hr incubation, the bacteria were centrifuged at 4000 x *g* for 20 min. Cells were resuspended in bacterial lysis buffer (75 mM Tris-HCl, pH 8; 0.25 M sucrose; 0.25 mM EDTA; 0.02 mg/ml lysozyme), and incubated 20 min on ice. The cells were repelleted at 3000 × *g*, resuspended in ice cold buffer A (35 mM potassium phosphate; 2.5 mM $MgCl_2$; 10 mM 2-mercaptoethanol, pH 7.4) and disrupted over ice by 4 cycles of 10 second sonication with a Branson model 350 sonifier fitted with a microtip. Cellular debris was removed by centrifugation at 100,000 × *g* for 30 min and the supernatant fraction diluted to a final concentration of 2.5 mg/ml with ice cold buffer A. The expression products were purified in one step by amylose affinity chromatography. The diluted supernatant was passed over a 2.5 × 10 cm column containing amylose resin and washed with 10 column volumes of cold buffer A.

Fusion protein was eluted with buffer A containing 10 mM maltose and the eluent concentrated in an Amicon Centriprep™ 30 concentrator. The maltose binding protein was cleaved from the expressed DPD using factor Xa according to manufacturer's instructions. Generation of the pMal-c2/bovine DPD construct resulted in the incorporation of a short segment of polylinker from the pCRII® vector (CTGGAATTCGGCTT) (SEQ ID NO:21) to the 5' end of bovine DPD cDNA. Following cleavage with factor Xa, the additional polylinker region on the cDNA and the use of the *Eco*RI cloning site in the pMal-c2 vector resulted in the addition of six amino acids (Ile, Ser, Glu, Phe, Gly, and Leu) on the N-terminal of the expressed DPD. Immunoblot analysis of expressed products was performed as previously described (Lu *et al.*, 1993).

10. DNA Sequencing

The bovine liver DPD cDNA was subjected to double-stranded sequencing by the dideoxynucleotide chain termination method using Sequenase® 2.0 and [α-^{35}S]dATP to label the newly synthesized strands. The ^{35}S-labeled products were resolved on 6% polyacrylamide-urea gels. The complete cDNA sequence was obtained by using commercially available or custom-made primers derived from cloned sequences. Sequence gels were read manually and analyzed using MacVector 4.1 Sequence Analysis software (IBI, New Haven, CT). Sequence analysis was repeated three times in each direction.

B. RESULTS

1. Generation of Full-length Bovine Liver DPD cDNA

The strategy described in FIG. 1A, FIG. 1B, FIG. 1C, and FIG. 1D yielded four cDNA fragments (65, 2360, 2076, 237 base pairs). Each fragment was shown to be part of the full-length bovine liver DPD cDNA (FIG. 2) by identification of peptide sequence derived from purified bovine liver DPD. The 65-bp fragment which was used as the starting point for this study contained amplified nucleotide sequence, which when translated, coded for a 9 amino-acid peptide (DALELNLSC) (SEQ ID NO:5) previously reported (Porter *et al.*, 1992a; 1992b) from a tryptic digest of purified bovine liver DPD. The 2360-bp fragment, extending to the 3' end of the cDNA, was verified as bovine liver DPD cDNA by the identification of nucleotide sequence which, when translated, coded for a 13 amino acid peptide (GLKADGTPWPAVG) (SEQ ID NO:22) isolated from CNBr digests of the purified enzyme. The 2076-bp fragment, extending to the 5' end of the cDNA, was verified as bovine liver DPD cDNA by the identification of amplified nucleotide sequence which when translated coded for a six amino acid peptide (SILALN) (SEQ ID NO:23) that was isolated from N-terminal peptide sequencing of purified bovine liver DPD. Lastly, the 237 base pair fragment, extending to the initiating ATG, was verified as bovine liver DPD cDNA by the identification of nucleotide sequence which when translated coded for the entire sixteen amino acid peptide sequence (VLSKDVADIESILALN) (SEQ ID NO:14).

Analysis of the sequence of the four larger fragments demonstrated overlapping regions with 100% sequence identity, which enabled alignment and assembly of the full-length cDNA as is shown in FIG. 2. The full-length clone (4414 base pairs) was generated by using restriction sites (*Bam*HI for the 2360 and 2076; *Msc*I for the 2076 and 237

base pair PCR™ products) common only in the overlapping regions to ligate the three fragments together (FIG. 2).

2. Sequence Analysis of Bovine Liver DPD cDNA

The nucleotide sequence and deduced amino acid sequence of the full-length bovine liver DPD cDNA are shown in FIG. 6A, FIG. 6B, and FIG. 6C. The start codon (shown in bold face) has the canonical flanking sequence for a translational start site with the customary GCC at position -3 to -1, and the standard G at position 4 (Kozak, 1991). The complete DPD cDNA sequence is 4414-bp long, contains a 74 nucleotide 5'-nontranslated region, and an open-reading frame of 3075 bases. The termination codon (TAA) is followed by 1273 nucleotides of the 3' non-translated region, including a 16-base poly $(A)^+$ tract. The open reading frame codes for a protein with a predicted molecular mass of 111,688 daltons. By comparison, purified DPD has an observed molecular mass of 108±3 kDa on SDS-PAGE (Lu *et al.*, 1993).

The deduced amino acid sequence was examined, utilizing the MacVector 4.1 Sequence Analysis software. Several protein motifs were identified in the translated sequence. These include a GDP/GTP binding site at position 1060, a 4-Fe/4-S binding site at position 1010 and a cAMP phosphorylation site at position 782 (Gilman, 1987; Otaka and Ooi, 1989).

3. Homology to Other Sequences

The nucleotide sequence for bovine liver DPD was compared to other sequences in the GenBank data base and confirmed as a new and unique sequence. Comparison of the deduced primary protein sequences indicated several sequences which share small regions of similarity. These sequences included three enzymes which contain either flavin-(dihydroorotate dehydrogenase and thioredoxin reductase) or NADPH-(glutamate synthase and thioredoxin reductase) binding domains (Quinn *et al.*, 1991; Russel and Model, 1988; Oliver *et al.*, 1987).

Dihydroorotate dehydrogenase demonstrated 40% identity over the 312 amino acids that their sequences overlapped. Regions of similarity were identified on a computer-generated dot-matrix plot. Those matches occurring within regions reported

(Quinn *et al.*, 1991) to be involved in flavin binding (corresponding to bovine DPD amino acid residues 787-792) are indicated.

Thioredoxin reductase demonstrated less homology to the bovine DPD sequence (37% identity) over a region of 333 overlapping amino acids, but conserved regions are apparent. These sites correspond to the reported FAD and NADPH binding sites for thioredoxin reductase (corresponding to bovine DPD amino acid residues 187-204 and 332-348, respectively) (Russel and Model, 1988). Glutamate synthase does not have a well-defined NADPH binding site (Oliver *et al.*, 1987) but the amino acid sequence demonstrated a 38% identity with a total overlap of 488 amino acids.

4. Northern Blot Analysis

The full-length bovine liver DPD cDNA was used as the probe in Northern analysis to determine the size and number of messages in both total and poly(A)$^+$ RNA from bovine liver (FIG. 3). With both types of RNA, a single band was observed with a size of about 4400 nucleotides. These results suggest that the complete cDNA has been isolated and that there is only a single gene transcript encoding bovine liver DPD.

5. *In Vitro* Transcription and Translation of Bovine Liver cDNA

In vitro transcription and translation were used to verify that the cloned cDNA translated a protein equivalent in size to bovine liver DPD. This procedure was performed prior to bacterial expression of the cDNA to confirm that the open reading frame contained no errant stop codons. Resolution of the labeled product by SDS-PAGE showed a single specific protein product with a molecular mass of approximately 108 kDa (FIG. 4). The data indicate that the cloned cDNA encodes a protein identical in size to purified bovine liver DPD and any post-translational modifications of the enzyme do not alter its mobility during SDS-PAGE.

6. Expression of Bovine Liver DPD cDNA in *E. coli*

The cloned DPD cDNA was inserted into the pMAL-c2 vector downstream from the *malE*gene, which encodes maltose-binding protein (MBP). This resulted in the expression of a MBP-DPD fusion protein containing amino acid residues 7-1025. Following cleavage of the fusion protein with Factor Xa (Nagai and Thogerson, 1984; Nagai and Thogerson, 1987), immunoblot analysis of expression products revealed a

protein that migrated at approximately 108 kDa and was recognized by the anti-human DPD polyclonal antibody (FIG. 5). Cytosol from control cells transfected with only the vector did not contain this immunoreactive protein. These data correlate with those obtained in the *in vitro* translation studies demonstrating that the translated product of the cDNA, the expressed protein, migrates at a molecular mass identical to that of purified, immunoreactive bovine liver DPD.

C. DISCUSSION

Based on the amino acid sequence of peptides derived from purified bovine liver DPD, the full-length cDNA was cloned, sequenced, and expressed in a bacterial cell line. Comparison to other sequences in the GenBank database verified that this is a unique sequence. The conclusion that the cDNA clone contained the entire coding region of bovine liver DPD is based on the following observations:

(1) the open-reading frame codes for a protein consisting of 1025 amino acids (molecular mass 111, 688 daltons) corresponding to that of purified enzyme (the active form of the enzyme is a homodimer, made up of two 108 kDa subunits (Lu *et al.*, 1993);

(2) the deduced amino acid sequence of the cloned cDNA contained all three amino acid sequences determined from purified enzyme;

(3) Northern blot analysis of bovine liver RNA detected a single band of appropriate length corresponding to the full-length cDNA;

(4) bacterial expression of the DPD cDNA generated a protein which comigrated with purified bovine liver DPD during SDS-PAGE, and when immunoblotted, reacted with a specific polyclonal rabbit anti-DPD antibody; and

(5) analysis of the bovine liver cDNA suggests the presence of prosthetic groups (FAD, NADPH and 4 Fe/4 S binding sites) known to be present on the purified protein.

A partial DPD cDNA (2300 base pairs long) was isolated from a λgt11 bovine liver cDNA library (Clontech) and found to have 100% sequence identity with the full-length bovine cDNA presented in this study. The partial clone began at nucleotide 425 and extended through to nucleotide 2765. Examination of the sequence following this region (nucleotides 2766-2776) shows a short poly-A tract that could have served as a

start site for first strand cDNA synthesis during construction of the library. This DPD cDNA fragment was isolated from the bovine liver cDNA library as a single clone and codes for most of the open reading frame of bovine liver cDNA. When translated, the amino acid sequence includes both the CNBr and tryptic peptide sequences derived from purified bovine liver DPD. Isolation of this partial DPD cDNA provides further evidence that this newly described full-length cDNA, which is a composite of three separate cDNA fragments, codes for the correct enzyme.

Database searches for amino acid sequences, similar to DPD, identified dihydroorotate dehydrogenase, thioredoxin reductase, and glutamate synthase with a partial amino acid sequence identity of 40, 37, and 38%, respectively. While these values are too low to support a common ancestry for these proteins (Doolittle, 1981), they do contain certain functional similarities to DPD. Dihydroorotate dehydrogenase is a flavoprotein (using FAD as a cofactor) which catalyzes the fourth step in pyrimidine biosynthesis (Quinn *et al.*, 1991). In addition, both thioredoxin reductase and glutamate synthase use NADPH as a cofactor (Russel and Model, 1988; Oliver *et al.*, 1987). Attempts to align these sequences to reveal structural motifs identifying common prosthetic groups demonstrated highly conserved areas which could represent the FAD and NADPH binding sites in bovine liver cDNA. Because these binding domains are likely to be in close proximity and probably have elements in common, it was not possible to resolve these two regions by using this comparison technique. Further analysis of the translated bovine DPD cDNA sequence revealed one 4 Fe/4 S binding site, a GDP/GTP binding site and a cAMP and cGMP dependent protein kinase phosphorylation site. The iron sulfur binding site is consistent with data obtained from purified DPD (Shiotani and Weber, 1981; Lu *et al.*, 1992).

Biological activity of DPD expressed using the prokaryotic pMAL vector was evaluated for both DPD/pMAL constructs and for vector controls. While only the pMAL construct containing the bovine DPD cDNA generated immunoreactive enzyme, neither sample contained significant DPD activity. This suggests the possibility that expression of the fusion protein could alter the folding of DPD and generate inactive enzyme. An

expression system such as baculovirus, which provides post-translational processing, might be more appropriate to generate enzymatically active DPD.

The availability of the complete nucleotide sequence of bovine liver DPD (together with the completely translated amino acid sequence) should permit further studies including: elucidation of tertiary structure; binding with known cofactors; and specific interactions with inactivators of this enzyme. Furthermore, the availability of the bovine liver DPD cDNA should allow the isolation of the full-length cDNA from other species including human. This in turn should provide insight into the molecular basis of the altered DPD activity observed with the inherited (pharmacogenetic) disorder (Diasio *et al.*, 1988; Lu *et al.*, 1993).

EXAMPLE 3

CLONING AND CHARACTERIZATION OF HUMAN LYMPHOCYTE DPD GENE

In this example, the human lymphocyte DPD cDNA was isolated and the DNA and polypeptide sequences were determined.

A. MATERIALS AND METHODS

1. Materials

The human lymphocyte λgt10 cDNA library was obtained from Clontech (Palo Alto, CA). Histopaque was purchased from the Sigma Chemical Co. (St. Louis, MO). The TA cloning kit was purchased from Invitrogen (San Diego, CA). Restriction enzymes and DNA-modifying enzymes were from New England Biolabs (Beverly, MA). The random primer labeling kit and microspin columns were from Pharmacia LKB Biotechnology Inc (Piscataway, NJ). The TNT coupled *in vitro* transcription and translation system, plasmid and lambda purification kits were purchased from Promega (Madison, WI). Specific oligonucleotides were synthesized by Midland Certified Reagent Co. (Midland, TX). The GeneAmp® PCR™ kit was obtained from Perkin-Elmer. The hybridization membranes were obtained from Schleicher & Schuell (Keene, NH). [α-^{32}P]dCTP (3000 Ci/mmol), [α-^{32}S]dATP (3000 Ci/mmol) and [^{35}S]methionine (1000 Ci/mmol) were purchased from Amersham (Arlington Heights, IL). All other reagents were of molecular biology grade.

2. First-Strand cDNA Synthesis

Bovine liver was obtained directly from a local slaughter house and snap frozen. Total RNA was isolated using RNAzol® B (Biotecx, Houston, TX) following the manufacturers instructions. Purification of poly(A)+ RNA was performed using an Oligotex-dT mRNA kit (Qiagen) according to the manufacturers instructions. Random primed cDNAs were prepared from 5 μg total RNA or poly(A)+ using a first-strand cDNA synthesis kit (Pharmacia). cDNA synthesis was performed in a 33 μl reaction volume following manufacturers instructions. The reaction mix was treated with 1 μl RNase H and incubated for 30 min at 37°C prior to PCR™ amplification.

3. cDNA Cloning of Human Lymphocyte DPD

Bovine liver DPD cDNA was cloned as described in Example 3. Human lymphocyte DPD cDNA was cloned by screening a λgt10 cDNA library grown in *E. coli* C600 Hfl bacteria on 150-mm diameter Petri dishes at a phage concentration of 40,000 plaques/plate. Duplicate nylon membranes (NYTRAN®, Schleicher & Schuell) from each plate were probed (Davis *et al.*, 1986) using labeled bovine liver DPD cDNA. The probe was labeled by random-primed synthesis (Feinberg and Vogelstein, 1984; Feinberg and Vogelstein, 1984) with [α-^{32}P]dCTP to a specific activity of 3.0×10^9 cpm/μg using a Pharmacia oligolabeling kit. The membranes were prehybridized for thirty minutes in QUIKHYB (Stratagene) and hybridized at 60°C for 2 hr in the same solution containing 100 μg/ml denatured salmon sperm DNA and the ^{32}P-labeled probe (2.5×10^6 cpm/ml). The membranes were washed twice in 2X SSC/0.1 % SDS for 15 min at room temperature followed by a 30 min wash at 58°C with 0.1X SSC/0.1% SDS and were exposed overnight to autoradiograph film at -70°C, with an intensifying screen.

Single positive plaque-forming units were isolated by cycles of dilution and rescreening. Phage DNA was purified using a lambda DNA purification kit (Promega) following manufacturer's instructions. The DPD cDNA insert was isolated from the λgt10 phage by digestion with *Eco*RI, purified on low melt agarose, and subcloned into the *Eco*RI site of pGEM-7zf® (Promega), for sequence analysis. The human cDNA was isolated as three overlapping fragments which were ligated together using overlapping restriction sites.

4. DNA Sequencing

DPD cDNA was subjected to double-stranded sequencing by the dideoxynucleotide chain termination method using Sequenase® 2.0 and [α-^{35}S]dATP (Sanger *et al.*, 1977). The ^{35}S-labeled products were resolved on 6% polyacrylamide-urea gels. The complete cDNA sequence was obtained by using commercially available or custom-made primers derived from cloned sequences. Sequence gels were read manually and analyzed using MacVector 4.1 Sequence Analysis software (IBI, New Haven, CT). Sequence analysis was repeated three times in each direction.

B. RESULTS and DISCUSSION

The nucleotide sequence and deduced amino acid sequence of full length human lymphocyte DPD cDNA are shown in FIG. 7A, FIG. 7B, FIG. 7C, and FIG. 7D. The ATG start codon (shown in bold face) has the canonical flanking sequence for a translational start site with the customary GCC at position -3 to -1, and the standard G at position 4 (Kozak, 1991). The complete human lymphocyte DPD cDNA sequence is 4356 base pairs long, contains a 48 nucleotide 5'-nontranslated region, and an open-reading frame of 3075 bases encoding a 1025 amino acid protein. The termination codon (TAA) is followed by 1230 nucleotides of 3' non-translated region containing at least two polyadenylation signal sequences (AAUAAA). Following elucidation of the sequence of human lymphocytic DPD cDNA, a data base search of GenBank revealed a recent publication reporting the cDNA and amino acid sequences of pig and human liver DPD (Yokota *et al.*, 1994). Comparison of the translated sequences revealed a single conservative amino acid difference between human lymphocyte (Asn) and human liver (Ser) DPD at position 910. A previous study had suggested that human lymphocytic DPD has a different isozyme from that of the liver (Naguib *et al.*, 1985). This was based on kinetic data which suggested the absence of allosterism in lymphocyte DPD and its presence in the liver enzyme. In addition, differences in K_m values for the liver enzyme versus that of the lymphocyte were observed. Furthermore, comparison of different tissues suggested that lymphocytes contain about five times more DPD activity compared to that found in liver (Naguib *et al.*, 1985).

EXAMPLE 4

IDENTIFICATION OF A FRAMESHIFT MUTATION IN THE DPD GENE IN A DPD-DEFICIENT PATIENT EXHIBITING FUra TOXICITY

In this example, specific primers were synthesized using the human lymphocyte DPD cDNA sequence as a template, and these primers were used to PCR™ amplify the cDNA encoding DPD from a DPD-deficient patient. Sequence analysis of the cDNA from this patient revealed the deletion of an adenosine nucleotide (_AA) compared to control (AAA) at codon 318. The resulting frameshift causes multiple incorrect codons and a premature stop in codon 335. *In vitro* transcription/translation of the cDNA of the deficient patient containing the frameshift resulted in a truncated protein with a molecular weight of approximately 40,000 daltons. PCR™ amplification of the patient's genomic DNA bordering the mutation (frameshift) demonstrated a mixed population containing both normal and mutated DNA demonstrating that the patient was heterozygous for this mutation, indicating that the mutation was the basis for DPD deficiency in the patient.

A. MATERIALS AND METHODS

1. RT-PCR™ of DPD cDNA from a DPD-Deficient Patient

Peripheral blood mononuclear cells were separated from 250 ml of whole blood collected from an individual with normal DPD activity (normal range 0.425 + 0.124 nmoles/min/mg protein) and a patient with partial DPD deficiency (less than 0.182 nmoles/min/mg protein) as previously described (Lu *et al.*, 1993). Total RNA was isolated using RNAzol B (Biotecx, Houston, TX) and random primed cDNAs were prepared using a first-strand cDNA synthesis kit (Pharmacia) as described above. PCR™ amplification was carried out under the conditions described above using two specific oligonucleotide primers which border the open-reading frame and include the initiating ATG in the sense primer (sense 5'-TGTAGGCACTGCCATGGCCCCTGTG-3') (SEQ ID NO:25) and the stop codon TAA in the antisense primer (antisense 5'-TTCACAAATCACCTTAACACACC-3') (SEQ ID NO:26); these primers correspond to positions 36-60 and 3117-3139, respectively, of the DPD cDNA sequence. The 3104 base pair PCR™ product containing the 3075 base pair open-reading frame was purified by electrophoresis in low melting point agarose and subcloned directly into the pCRII® vector (Invitrogen).

2. *In vitro* Transcription and Translation of Human DPD cDNA

Human DPD cDNA clones (3104 base pairs) from normal and the DPD deficient patient were subcloned in the pCRII® vector downstream from an Sp6 promoter site. *In vitro* transcription and translation was conducted with the TNT' Sp6 coupled reticulocyte lysate system (Promega) using either [^{35}S]methionine for labeling of the synthesized proteins or unlabeled amino acid for western blot analysis. The translated products were resolved by SDS-PAGE on a 5 to 15% gradient polyacrylamide gel. The gels using [^{35}S]methionine for labeling of the synthesized proteins were vacuum-dried at 65°C and exposed to autoradiograph film for 1 hour. Gels containing unlabeled proteins were transferred to nitrocellulose and immunoblotted as described Lu *et al.*, 1992).

3. Northern Blot Analysis

Total and poly(A)+ RNA were prepared by the methods described above for cDNA synthesis, resolved by formaldehyde/agarose gel electrophoresis, and transferred onto NYTRAN nylon membranes. Radiolabeled probe (specific activity = 1×10^{11} cpm/μg) was prepared with a Pharmacia Oligolabelling Kit using full-length bovine liver or human lymphocyte cDNA as the template. The filters were UV-cross-linked, prehybridized for 30 min, and then hybridized for 2 hours in Quikhyb®. The blots were washed under high stringency conditions and developed with autoradiograph film overnight at -70°C.

4. Genomic DNA Preparation

Genomic DNA was prepared from human peripheral blood mononuclear cells from both the normal and the DPD deficient patient as previously described (Sambrook *et al.*, 1989). 200 ng of genomic DNA was used as a template in the PCR™ reaction. Following denaturation at 95°C for two minutes, 35 cycles of denaturation at 95°C (1 min), annealing at 55°C (1 min), and extension at 72°C (1 min) were performed. The primers used (sense 5'-TTGGTGGTTTAAGTACTTCTGAAATTCC-3') (SEQ ID NO:27) and (antisense 5'-CTTGCTCTGTCCGAACAAACTGC ATAGCA-3') (SEQ ID NO:28), corresponded to positions 716-743 and 1260-1288, respectively, of the DPD cDNA sequence. These primers were designed to amplify a 573 bp fragment from the exon in genomic DNA that contained the mutation. Following amplification, the PCR™ product was purified on 1.5% low melting point agarose gel and subcloned directly into the pCRII® vector (Invitrogen).

B. RESULTS and DISCUSSION

1. Characterization of the Molecular Defect in a Patient Exhibiting FUra Toxicity Secondary to DPD Deficiency

Western blot analysis of cytosol prepared from the DPD deficient patient's lymphocytes revealed a faint band comigrating with purified DPD. For quantitative analysis of DPD protein, different amounts of cytosol from the DPD deficient patient were subjected to western blot densitometric analysis (FIG. 8). These data suggested that the patient's cytosol contained approximately 10-fold less DPD as compared to a positive control prepared from the individual having normal DPD activity.

The availability of the cDNA for human lymphocyte DPD has permitted further studies of this patient (Johnson *et al.*, 1995). Northern blot analysis was utilized to determine the size and number of messages from the individual with normal DPD activity and the DPD deficient patient. Human lymphocyte DPD cDNA was used as a probe and demonstrated a single band with a size of approximately 4400 nucleotides for both the normal and DPD deficient patient (FIG. 9). These results suggest that there is only a single gene transcript coding for DPD. In addition, these data eliminated the possibility that a large insertion or deletion occurred in the message of the deficient patient

(possibly caused by a splicing problem). The intensity of the band for the deficient patient corresponded to that of the control RNA suggesting that the message was efficiently transcribed and stable.

Having established that the patient transcribes poly(A)+ RNA coding for DPD, RT-PCR™ was used to clone the cDNA coding for DPD from both the normal and DPD deficient patient. Following first strand cDNA synthesis, PCR™ amplification was carried out using primers which border the open-reading frame and include the initiating ATG in the sense primer (sense 5'-TGTAGGCACTGCCATGGCCCCTGTG-3') (SEQ ID NO:25) and the stop codon TAA in the antisense primer (antisense 5'-TTCACAAATCACCTTAACACACC-3') (SEQ ID NO:26). The 3104 base pair PCR™ product, containing the 3075 base pair open-reading frame, was subcloned directly into the pCRII® vector. *In vitro* transcription/translation was used to verify that the cloned cDNA from the deficient patient translated a protein equivalent in size and immunoreactivity to that generated by the cDNA from the individual with normal DPD activity. Reactions were performed using [^{35}S]-methionine for labeling the synthesized proteins. Analysis of the labeled proteins by SDS-PAGE revealed that two out of the ten subclones tested generated a truncated protein with a molecular weight of approximately 40,000 daltons as compared to 108,000 daltons for the control (FIG. 10A and FIG. 10B). Translation was repeated in the presence of unlabeled amino acids and the products examined for immunoreactivity to a specific rabbit anti-human DPD polyclonal antibody by western blot analysis. As shown in FIG. 10A and FIG. 10B, the 40,000 dalton protein band was immunoreactive against the specific anti-DPD polyclonal antibody. The formation of a truncated form of DPD suggested either an insertion, deletion, or nonsense mutation in the open reading frame of the DPD deficient patient's cDNA. Based on the size of the truncated DPD, specific oligonucleotide primers were used to sequence a 573 base pair fragment of the cDNA; a stop codon within this region would result in a protein ranging in size from 25,000-45,000 daltons. Sequence analysis of the cDNA from the DPD deficient patient revealed a single base pair deletion at position 1000 corresponding to codon 318 (FIG. 11). This deletion causes a frameshift that results in truncation of translation at codon 335 generating a 36,500 dalton protein.

2. DNA Sequence Analysis of DPD Gene in a DPD-Deficient Patient

Complete sequence analysis of the DPD deficient patient's cDNA also revealed an additional single nucleotide difference from that of control; A (control) to T (deficient) at position 2894. Translation of the cDNA demonstrated that this resulted in a nonconservative amino acid substitution (Asp to Val). Subsequent subcloning and sequence analysis of multiple PCR™ reactions flanking this region from a number of individuals having normal DPD activity demonstrated that this nucleotide substitution was common in the general population and may represent an allelic variant. In contrast, the adenosine deletion resulting in a frameshift was not found in any individuals having normal DPD activity but was identified exclusively in the DPD deficient patient's cDNA.

Since this deletion was initially identified in the cDNA of the DPD deficient patient (two out of the ten subclones), studies were undertaken with genomic DNA to confirm that this patient was heterozygous for this mutation. Primers were designed based on the cDNA sequence to amplify a 573 base pair DNA fragment from the exon containing the sequence of interest (FIG. 7A, FIG. 7B, FIG. 7C, and FIG. 7D). Sequence analysis of several clones (from multiple PCR™ reactions) from the deficient patient indicated the presence of two different alleles (one of these containing the deletion, the other identical to normal), present in approximately equal amounts. The identification of both the normal and mutant allele (adenosine deletion) in the genomic DNA confirm that this patient is heterozygous for the single base deletion.

In summary, the gene and the poly(A)+ RNA encoding the DPD protein in this patient contains an adenosine deletion that causes a frameshift resulting in truncation of translation at codon 335 generating a 36,500 dalton protein. Analysis of the patient's genomic DNA has demonstrated that this patient is heterozygous for this mutation. This represents the first molecular characterization of a DPD deficient patient, and provides an explanation for reduced DPD activity. This frameshift has also been identified in an additional unrelated DPD deficient patient who also exhibited severe FUra toxicity.

EXAMPLE 5

MONOCLONAL ANTIBODY GENERATION

Means for preparing and characterizing antibodies are well known in the art (See, *e.g.*, Howell and Lane, 1988). The methods for generating monoclonal antibodies (mAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic composition in accordance with the present invention and collecting antisera from that immunized animal. A wide range of animal species can be used for the production of antisera. Typically the animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, *m*-maleimidobencoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

As is also well known in the art, the immunogenicity of a particular immunogen composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Exemplary and preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis)*, incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized

animal at various points following immunization. A second, booster, injection may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal can be bled and the serum isolated and stored, and/or the animal can be used to generate mAbs.

mAbs may be readily prepared through use of well-known techniques, such as those exemplified in U.S. Patent 4,196,265, incorporated herein by reference. Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, *e.g.*, a purified or partially purified LTBP-3 protein, polypeptide or peptide. The immunizing composition is administered in a manner effective to stimulate antibody producing cells. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986, pp. 60-61), but mice are preferred, with the BALB/c mouse being most preferred as this is most routinely used and generally gives a higher percentage of stable fusions.

Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B-cells), are selected for use in the mAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible. Often, a panel of animals will have been immunized and the spleen of animal with the highest antibody titer will be removed and the spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately 5×10^7 to 2×10^8 lymphocytes.

The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render then incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, pp. 65-66, 1986; Campbell, pp. 75-83, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse murine myeloma SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 ratio, though the ratio may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described (Kohler and Milstein, 1975; 1976), and those using polyethylene glycol (PEG), such as 37% (v/v) PEG (Gefter *et al.*, 1977). The use of electrically induced fusion methods is also appropriate (Goding pp. 71-74, 1986).

Fusion procedures usually produce viable hybrids at low frequencies, about 1×10^{-6} to 1×10^{-8}. However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is

supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, *e.g.*, hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B-cells can operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B-cells.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immuno-binding assays, and the like.

The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones can then be propagated indefinitely to provide mAbs. The cell lines may be exploited for mAb production in two basic ways. A sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide mAbs in high concentration. The individual cell lines could also be cultured *in vitro*, where the mAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations. mAbs produced by either means may be further purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography.

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the

compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the composition, methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following literature citations as well as those cited above are incorporated in pertinent part by reference herein for the reasons cited in the above text:

Allen and Choun, "Large unilamellar liposomes with low uptake into the reticuloendothelial system," *FEBS Lett.*, 223:42-46, 1987.

Allen, *In: Laboratory Techniques in Biochemistry and Molecular Biology, Volume on Sequencing of Protein and Peptides*, (Burdon, R.H., and van Knippenber, P.H., eds) pp. 58-59, Elsevier Science Publishing Co., Inc., New York, 1989.

Altschul *et al.*, "Basic local alignment search tool," *J. Mol. Biol.*, 215:403-410, 1990.

Ausebel *et al.*, *Current Protocols in Molecular Biology*, Wiley, New York, 1987.

Baccanari *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:11064-11068, 1993.

Bauer *et al.*, *J. Biol. Chem.*, 257:3593-3597, 1982.

Bischoff *et al.*, *Eur. J. Immunol.*, 23:761-767, 1993.

Boshart *et al.*, *Cell*, 66:849-859, 1991.

Brashekka *et al.*, *J. Neurol. Sci.*, 78:71-77, 1987.

Brocksedt *et al.*, *J. Inherited Metab. Dis.*, 13:121-124, 1990.

Broxmeyer *et al.*, *Blood*, 76:1110-1116, 1990.

Broxmeyer *et al.*, *J. Immunol.*, 150:3448-3458, 1993.

Campbell, "Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology," Vol. 13, Burden and Von Knippenberg, Eds. pp. 75-83, Elsevier, Amsterdam, 1984.

Campbell, *In: Monoclonal Antibody Technology*, Laboratory Techniques in Biochemistry and Molecular Biology Vol. 13, Burden and Von Knippenberg, eds. pp. 75-83, Amsterdam, Elseview, 1984.

Capecchi, "High efficiency transformation by direct microinjection of DNA into cultured mammalian cells," *Cell* 22(2):479-488, 1980.

Caput *et al.*, *Proc. Natl. Acad. Sci. USA*, 83:1670-1674, 1986.

Chou and Fasman, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 47:45-148, 1978a.

Chou and Fasman, *Ann. Rev. Biochem.*, 47:251-276, 1978b.

Chou and Fasman, *Biochemistry*, 13(2):211-222, 1974b.

Chou and Fasman, *Biochemistry*, 13(2):222-245, 1974a.

Chou and Fasman, *Biophys. J.*, 26:367-384, 1979.

Church *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:1991-1995.

Clapp, "Somatic gene theraphy into hematopoietic cells. Current status and future implications," *Clin. Perinatol.* 20(1):155-168, 1993.

Cleland, *Methods Enzymol.*, 63:103-138, 1979.

Couvreur, "Polyalkyleyanoacrylates as colloidal drug carriers," *Crit. Rev. Ther. Drug Carrier Syst.*, 5:1-20, 1988.

Couvreur *et al.*, "Nanocapsules, a new lysosomotropic carrier," *FEBS Lett.*, 84:323-326, 1977.

Cox *et al.*, *J. Virol.* 67(9):5664-5667, 1993.

Curiel *et al.*, "Adenovirus enhancement of transferrin-polylysine-mediated gene delivery," *Proc. Natl. Acad. Sci. USA* 88(19):8850-8854, 1991.

Curiel *et al.*, "P.C. high-efficiency gene transfer mediated by adenovirus coupled to DNA-polylysine complexes," *Hum. Gen. Ther.* 3(2):147-154, 1992.

Daher *et al.*, *Biochem. Pharmacol.*, 41:1887-1893, 1991.

Daher *et al.*, *Drug Metab. Dispos.*, 19:285-287, 1991.

Daher *et al.*, *Pharmac. Ther.*, 48:189-222, 1990.

Davatelis *et al.*, *J. Exp. Med.*, 167:1939-1944, 1988.

Davis *et al.*, *Basic Methods in Molecular Biology*, Elsevier Press, New York, NY, 1986.

Desiderio and Campbell, "Liposome-encapsulated cephalotin in the treatment of experimental murine-salmonellosis," *J. Reticuloendothel. Soc.*, 34:279-287, 1983.

Diasio and Harris, *Clin. Pharmacokinet.*, 16:215-237, 1989.

Diasio and Lu, *J. Clin. Oncol.*, 12:2239-2242, 1994.

Diasio *et al.*, *J. Clin. Invest.*, 81:47-51, 1988.

Diasio *et al.*, The 8th International Symposium Purine & Pyrimidine, Bloomington, IN, May 22-27, 1994

Doolittle, *Science*, 214:149-159, 1981.

Eglitis and Anderson, "Retroviral vectors for introduction of genes into mammalian cells," *Biotechniques* 6(7):608-614, 1988.

Eglitis *et al.*, "Retroviral-mediated gene transfer into hemopoietic cells," *Avd. Exp. Med. Biol.* 241:19-27, 1988.

Faeder and Siegel, *Anal. Biochem.*, 53:332-336, 1973.

Feinberg and Vogelstein, *Anal. Biochem.*, 137:266, 1984.

Feinberg and Vogelstein, *Anal. Biochem.*, 132:6, 1983.

Fritzson, *J. Biol. Chem.*, 235:719-725, 1960.

Frohman *et al., PCR Protocols: A Guide to Methods and Applications.* (Innis, M. A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds), Academic Press, San Diego, CA., 1990.

Frohman *et al., Proc. Natl. Acad. Sci. USA*, 85:8998-9002, 1988.

Frohman, *PCR Protocols: A Guide to Methods and Applications.* (Innis, M. A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds), Academic Press, San Diego, CA, 1990.

Fromm *et al.*, "Expression of genes transferred into monocot and dicot plant cells by electroporation," *Proc. Natl. Acad. Sci. USA* 82(17):5824-5828, 1985.

Fujimoto *et al., J. Nutr. Sci. Vitaminol.*, 37:89-98, 1990.

Fynan *et al.*, "DNA vaccines: protective immunizations by parenteral, mucosal, and gene gun inoculations," *Proc. Natl. Acad. Sci. USA* 90(24):11478-11482, 1993.

Fynan *et al.*, "DNA vaccines: Protective immunizations by parenteral, mucosal, and gene-gun inoculations," *Proc. Natl. Acad. Sci. USA*, 90:11478-11482, 1993.

Gaastra, *In: Methods in Molecular Biology, Volume I. Proteins* (Walker, J.M., ed), Humana Press, Clifton, pp. 349-355, NJ., 1984.

Gabizon and Papahadjopoulos, "Liposomes formulations with prolonged circulation time in blood and enhanced uptake by tumors," *Proc. Natl. Acad. Sci. USA*, 85:6949-6953, 1988.

Gefter *et al., Somatic Cell Genet.*, 3:231-236, 1977.

Gefter *et al., Somatic Cell Genet.*, 3:231-236, 1977.

Gilman, *Ann Rev Biochem* 56:615649, 1987.

Gilman *et al., Ann Rev Biochem*, 56:615-649, 1987.

Goding, "Monoclonal Antibodies: Principles and Practice," pp. 60-74. 2nd Edition, Academic Press, Orlando, FL, 1986.

Goding, *In: Monoclonal Antibodies: Principles and Practice*, 2d ed., Orlando, Fla., Academic Press, pp. 65-74, 1986.

Goding, *In: Monoclonal Antibodies: Principles and Practice*, 2d ed., Orlando, Fla., Academic Press, pp. 60-61, 65-66, 71-74, 1986.

Goedde *et al.*, *Hoppe-Seyler's Z. Physiol. Chem.*, 351:945-951, 1970.

Goeddel *et al.*, *Nature*, 281:544, 1979.

Goeddel *et al.*, *Nucleic Acids Res.*, 8:4057, 1980.

Graham and van der Eb, "Transformation of rat cells by DNA of human adenovirus 5," *Virology* 54(2):536-539, 1973.

Grisolia and Carrdoso, *Biochim. Biophys. Acta*, 25:430-431, 1957.

Grynkiewicz *et al.*, *J. Biol. Chem.*, 260:340, 1985.

Guan *et al.*, *Gene*, 67:21-30, 1987.

Hannum *et al.*, *Nature*, 343:336, 1990.

Harlow and Lane, "Antibodies: a Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1988.

Harris *et al.*, *Biochem. Pharmacol*, 37:4759-4762, 1988.

Harris *et al.*, *Cancer*, 68:499-501, 1991.

Harris *et al.*, *Cancer Res.*, 50:197-201, 1990.

Henry-Michelland *et al.*, "Attachment of antibiotics to nanoparticles; Preparation, drug-release and antimicrobial activity *in vitro*, *Int. J. Pharm.*, 35:121-127, 1987.

Ho *et al.*, *Anticancer Res.*, 6:781-784, 1986.

Holopainen and Konto, *Neurochem. Res.*, 11:207-215, 1986.

Horuk *et al.*, *Biochemistry*, 32:5733-5738, 1993.

Horuk *et al.*, *Immunol. Today*, 15;169, 1994.

Horuk *et al.*, *Science*, 261:1182-1184, 1993.

Howell and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988.

Iigo *et al.*, *Biochem. Pharmacol.*, 38:1885-1889, 1969.

Jameson and Wolf, *Compu. Appl. Biosci.*, 4(1):181-6, 1988.

Jameson and Wolf, *Compu. Appl. Biosci.*, 4(1):181-6, 1988.

Johnson *et al.*, *Proc. Amer. Assoc. Can. Res.* (Abst.) 2992 (In Press), 1995.

Johnston and Tang, "Gene gun transfection of animal cells and genetic immunization," *Methods Cell. Biol.* 43(A):353-365, 1994.

Kelvin *et al.*, *J. Leuk. Biol.*, 54:604, 1993.

Killary and Fournier, *Cell*, 38:523-534, 1984.

Kohler and Milstein, *Eur. J. Immunol.*, 6:511-519, 1976.

Kohler and Milstein, *Eur. J. Immunol.*, 6:511-519, 1976.

Kohler and Milstein, *Nature*, 256:495-497, 1975.

Kohler and Milstein, *Nature*, 256:495-497, 1975.

Kozak, *J. Biol. Chem.*, 266:19867-19870, 1991.

Kruys *et al.*, *Enzyme*, 44:193-202, 1993.

Kruys *et al.*, *Gene*, 72:191-200, 1988.

Kruys *et al.*, *J. Exp. Med.*, 177:1383-1390, 1993.

Kruys *et al.*, *Proc. Natl. Acad. Sci.*, 84:6030-6034, 1987.

Kruys *et al.*, *Science*, 245:852-855, 1989.

Krzyzek, "Genetic transformation of maize cells by electroporation of cells pretreated with pectin degrading enzymes," U.S. Patent 5,384,253, 01/24/95.

Kuby, "Immunology" 2nd Edition. W. H. Freeman & Company, New York, 1994.

Kuna *et al.*, *J. Immunol.*, 150:1932-1943, 1993.

Kyte and Doolittle, *J. Mol. Biol.*, 157:105-132, 1982.

Kyte and Doolittle, *J. Mol. Biol.* 157(1):105-132, 1982.

Laemmli, *Nature*, 227:680, 1970.

Lee *et al.*, *Science*, 239:1288-1291, 1988.

Lowry *et al.*, *J. Biol. Chem.*, 193:265-275, 1951.

Lu *et al.*, *Biochem. Pharm.*, 46:945-952, 1993b.

Lu *et al.*, *Cancer Res.*, 53:5433-5438, 1993a.

Lu *et al.*, *J. Biol. Chem.*, 267:171102-17109, 1992.

Lu *et al.*, "High efficiency retroviral mediated gene transduction into single isolated immature and replatable CD34(3+) hematopoietic stem/progenitor cells from human umbilical cord blood," *J. Exp. Med.* 178(6):2089-2096, 1993.

Lundquist *et al.*, "Fertile transgenic *Zea mays* plants -- comprising recombinant DNA encoding seed storage protein," WO/9110725 07/25/91.

Maloy *et al.*, "Microbial Genetics" 2nd Edition. Jones and Bartlett Publishers, Boston, MA, 1994.

Maloy, "Experimental Techniques in Bacterial Genetics", Jones and Bartlett Publishers, Boston, MA 1990.

Maniatis *et al.*, "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY., 1982.

Martin *et al.*, *Cancer Bull.*, 30:219-224, 1978.

Massey and Swoboda, *Biochem. Z.* 338:474-484, 1963.

Matsudaira *et al.*, *J. Biol. Chem.* 262:10035-10038, 1987.

Matthews *et al.*, *Arch. Biochem. Biophys.*, 293:254-263, 1992.

McColl *et al.*, *J. Immunol.*, 150:4550, 1993.

Merril *et al.*, *Science*, 211:1437-1438, 1981.

Myers *et al.*, *Neurochem. Res.*, 18:667-673, 1993.

Nagai and Thogersen, *Meth. Enzymol.*, 153:461-487, 1987.

Nagai and Thogersen, *Nature*, 309:810-812, 1984.

Naguib *et al.*, *Cancer Res.*, 45:5405-5412, 1985.

Nakamura *et al.*, Enzyme Immunoassays: Heterogeneous and Homogeneous Systems, Chapter 27, 1987.

Neote *et al.*, *Blood*, 84:44, 1994.

Neote *et al.*, *Cell*, 72:415-425, 1993.

Neote *et al.*, *J. Biol. Chem.*, 268:12247-12249, 1993.

Oliver *et al.*, *Gene*, 60:1-11, 1987.

Oppenheim *et al.*, *Ann. Rev. Immunol.*, 9:617, 1991.

Otaka and Ooi, *J. Mol. Evol.*, 29:246-254, 1989.

Pearson, *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:2444-2448, 1988.

Peppel *et al., J. Exp. Med.*, 174:1483, 1991; 1993.

Peppel *et al., J. Immunol.*, 151:5699, 1991; 1993.

Petit *et al., Proc. Natl. Acad. Sci. USA*, 83:2561-2565, 1986.

Podschun *et al., Eur. J. Biochem.*, 185:219-224, 1989.

Podschun *et al., J. Biol. Chem.*, 265:12966-12972, 1990.

Porter *et al., J. Biol. Chem.*, 266:19988-19994, 1992a.

Porter *et al., J. Biol. Chem.*, 267:5236-5242, 1992b.

Prokop and Bajpai, "Recombinant DNA Technology I" *Ann. N.Y. Acad. Sci.*, Vol. 646, 1991.

Quinn *et al., J. Biol. Chem.*, 266:9113-9127, 1991.

Rabinowitz, *Methods Enzymol.*, 53:275-277, 1978.

Ralph, *Lymphokine & Cytokine Res.*, 10:237-237, 1991.

Rot *et al., J. Exper. Med.*, 176:1489-1495, 1992.

Ruppert *et al., Cell*, 61:895-904, 1990.

Russel and Model, *J. Biol. Chem.*, 263:9015-9019, 1988.

Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual," 2nd Edition; Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

Sanger *et al., Proc. Natl. Acad. Sci. USA*, 74:5463-5467, 1977.

Sanno *et al., J. Biol. Chem.* 245:5668-5676, 1970.

Schall *et al., J. Exper. Med.*, 177:1821-1826, 1993.

Schall, *Cytokine*, 3:165-183, 1991.

Segal, "Biochemical Calculations" 2nd Edition. John Wiley and Sons, New York, 1976.

Shiotani and Weber, *J. Biol. Chem.*, 256:219-224, 1981.

Sommadossi *et al., J. Biol. Chem.*, 257:8171-8176, 1982.

Spoerel, *Methods Enzymol.* 152:588-597, 1987.

Tabor and Richardson, *Proc. Natl. Acad. Sci.*, 82:1074-1078, 1985.

Tang *et al., Nature*, 356:152-154, 1992.

Thomas *et al.*, "High yield purification of melanoma growth stimulatory activity," *Molec. Cell. Endocrin.*, 57: 69-76, 1988.

Towbin *et al., Proc. Natl. Acad. Sci. U.S.A.*, 76:4350-4354, 1979.

Traut and Loechel, *Biochemistry*, 23:2533-2539, 1984.

U.S. Patent 3,791,932.

U.S. Patent 3,949,064.

U.S. Patent 4,174,384.

U.S. Patent 4,196,265.

U.S. Patent 4,554,101.

U.S. Patent 4,578,770.

U.S. Patent 4,596,792.

U.S. Patent 4,599,230.

U.S. Patent 4,599,231.

U.S. Patent 4,601,903.

U.S. Patent 4,608,251.

U.S. Patent 4,683,195.

U.S. Patent 4,683,202.

U.S. Patent 4,965,188.

U.S. Patent 5,176,995.

Ulmer *et al.*, "Heterologous Protection Against Influenza by Injection of DNA Encoding a Viral Protein," *Science*, 259:1745-1749, 1993.

van Gennip *et al.*, *Adv. Exp. Med. Biol.*, 253A:111-118, 1989.

Wadman *et al.*, *J. Inherited Metab. Dis.*, 8(Suppl. 2):113-114, 1985.

Wagner *et al.*, "Coupling of adenovirus to transferrin-polylysine/DNA complexes greatly enhances receptor-mediated gene delivery and expression of transfected genes," *Proc. Natl. Acad. Sci. USA* 89(13):6099-6103, 1992.

Wang *et al.*, *J. Exp. Med.*, 177:699, 1993.

Wang *et al.*, *J. Immunol.*, 150:3022, 1993.

Wang *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:4156-4160, 1993.

Wasternack, *Biochem. Physiol. Pflanz.*, 173:371-403, 1978.

Wasternak, *Pharmac. Ther.*, 8:629-651, 1980.

Whitton *et al.*, *J. Virol.* 67:(1)348-352,1993.

Wolf *et al.*, *Compu. Appl. Biosci.*, 4(1):187-91, 1988.

Wolpe *et al., J. Exp. Med.*, 167:570-581, 1988.

Yokota *et al., J. Biol. Chem.*, 269:23192-23196, 1994.

Young and Davis, "Efficient Isolation of Genes by Using Antibody Probes," *Proc. Natl. Acad. Sci. USA*, 80:1194-1198, 1983.

Zafra *et al., Neurochem. Res.*, 9:695-707, 1984.

Zatloukal, L., Wagner, E., Cotten, M., Phillips, S., Plank, C., Steinlein, P., Curiel, D.T., and Birnstiel, M.L., "Transferrinfection: a highly efficient way to express gene constructs in eukaryotic cells," *Ann. N.Y. Acad. Sci.* 660:136-153, 1992.

Zawada *et al., Brain Res. Bull.*, 32:17-21, 1993.

Zhang *et al., Proc Am Assoc Cancer Res.*, 35:200, 1994.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: The UAB Research Foundation

(B) STREET: 701 South 20th Street, Suite 112 Administration Building

(C) CITY: Birmingham

(D) STATE: Alabama

(E) COUNTRY: USA

(F) POSTAL (ZIP) CODE: 35294-0111

(ii) INVENTORS: DIASIO, ROBERT B.
JOHNSON, MARTIN
ALBIN, NICOLAS
LU, ZHIHONG
ZHANG, RUIWEN

(iii) TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDRO[illegible]
AND METHODS OF USE

(iv) NUMBER OF SEQUENCES: 28

(v) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Arnold, White & Durkee

(B) STREET: P.O. Box 4433

(C) CITY: Houston

(D) STATE: Texas

(E) COUNTRY: United States of America

(F) ZIP: 77210

(vi) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vii) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: UNKNOWN

(B) FILING DATE: CONCURRENTLY HEREWITH

(C) CLASSIFICATION: UNKNOWN

(viii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/227,357

(B) FILING DATE: 13-APR-1994

(C) CLASSIFICATION: Unknown

(ix) ATTORNEY/AGENT INFORMATION:

(A) NAME: Wilson, Mark B.

(B) REGISTRATION NUMBER: 37,259

(C) REFERENCE/DOCKET NUMBER: UOAB025P--

(x) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (512) 418-3000

(B) TELEFAX: (713) 789-2679

(C) TELEX: 79-0924

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4414 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 68..3142

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACTTTCGCTG AAGCCTGAGG ACGCGGAAGG GTTCGTGGCA AGGAAACCCC AGGCTCTGGG 60

CATCGCC ATG GCC CCT GTG CTA AGC AAG GAC GTG GCG GAT ATC GAG AGT 109
Met Ala Pro Val Leu Ser Lys Asp Val Ala Asp Ile Glu Ser
1 5 10

ATC CTG GCT TTA AAT CCT CGA ACA CAG TCT CGT GCA ACT CTG CGT TCC 157
Ile Leu Ala Leu Asn Pro Arg Thr Gln Ser Arg Ala Thr Leu Arg Ser
15 20 25 30

ACT TTG GCC AAG AAA TTA GAC AAG AAA CAT TGG AAA AGA AAT CCT GAT 205
Thr Leu Ala Lys Lys Leu Asp Lys Lys His Trp Lys Arg Asn Pro Asp
35 40 45

AAG AAC TGC TTT AAT TGC GAG AAG CTG GAG AAT AAT TTT GAT GAC ATC 253
Lys Asn Cys Phe Asn Cys Glu Lys Leu Glu Asn Asn Phe Asp Asp Ile
50 55 60

AAG CAC ACG ACT CTT GGT GAG CGA GGA GCT CTC CGA GAA GCA ATG AGA 301
Lys His Thr Thr Leu Gly Glu Arg Gly Ala Leu Arg Glu Ala Met Arg
65 70 75

TGC CTG AAA TGT GCA GAT GCC CCC TGT CAG AAG AGC TGT CCA ACA AAT 349
Cys Leu Lys Cys Ala Asp Ala Pro Cys Gln Lys Ser Cys Pro Thr Asn
80 85 90

CTA GAT ATC AAA TCG TTC ATC ACA AGT ATC TCA AAC AAG AAC TAT TAT 397
Leu Asp Ile Lys Ser Phe Ile Thr Ser Ile Ser Asn Lys Asn Tyr Tyr

95 100 105 110

GGA GCT GCT AAG ATG ATA TTT TCT GAC AAC CCA CTT GGT CTG ACC TGT 445
Gly Ala Ala Lys Met Ile Phe Ser Asp Asn Pro Leu Gly Leu Thr Cys
115 120 125

GGA ATG GTA TGC CCC ACC TCT GAT CTT TGT GTA GGT GGA TGT AAT TTG 493
Gly Met Val Cys Pro Thr Ser Asp Leu Cys Val Gly Gly Cys Asn Leu
130 135 140

TAT GCC ACT GAA GAG GGA CCA ATT AAT ATT GGT GGA TTG CAG CAA TAT 541
Tyr Ala Thr Glu Glu Gly Pro Ile Asn Ile Gly Gly Leu Gln Gln Tyr
145 150 155

GCT ACT GAG GTA TTC AAA GCA ATG AAT ATC CCA CAA ATC AGG AAT CCT 589
Ala Thr Glu Val Phe Lys Ala Met Asn Ile Pro Gln Ile Arg Asn Pro
160 165 170

TCT CTG CCT CCC CCA GAA AAA ATG CCT GAA GCC TAT TCT GCA AAG ATT 637
Ser Leu Pro Pro Pro Glu Lys Met Pro Glu Ala Tyr Ser Ala Lys Ile
175 180 185 190

GCT CTT TTG GGT GCT GGG CCT GCA AGT ATA AGT TGT GCT TCG TTC TTG 685
Ala Leu Leu Gly Ala Gly Pro Ala Ser Ile Ser Cys Ala Ser Phe Leu
195 200 205

GCT CGA TTA GGC TAC AAT GAC ATC ACT ATA TTT GAA AAA CAA GAA TAC 733
Ala Arg Leu Gly Tyr Asn Asp Ile Thr Ile Phe Glu Lys Gln Glu Tyr
210 215 220

GTT GGT GGT ATA AGT ACT TCT GAA ATC CCT CAG TTC CGG CTG CCA TAT 781
Val Gly Gly Ile Ser Thr Ser Glu Ile Pro Gln Phe Arg Leu Pro Tyr
225 230 235

GAT GTA GTA AAT TTT GAG ATT GAG CTT ATG AAG GAC CTT GGT GTA AAG 829
Asp Val Val Asn Phe Glu Ile Glu Leu Met Lys Asp Leu Gly Val Lys

240 245 250

ATA ATT TGT GGT AAA AGC CTT TCA GTG AAT GAC ATT ACT CTT AGT ACT 877
Ile Ile Cys Gly Lys Ser Leu Ser Val Asn Asp Ile Thr Leu Ser Thr
255 260 265 270

TTG AAA GAA GAA GGG TAC AAA GCT GCT TTC ATT GGG ATA GGT TTG CCA 925
Leu Lys Glu Glu Gly Tyr Lys Ala Ala Phe Ile Gly Ile Gly Leu Pro
275 280 285

GAA CCC AAG AAG GAT CAC ATC TTC CAA GGC CTG ACA CAG GAC CAG GGG 973
Glu Pro Lys Lys Asp His Ile Phe Gln Gly Leu Thr Gln Asp Gln Gly
290 295 300

TTT TAC ACT TCC AAA GAC TTT CTG CCT CTT GTA GCC AAA AGC AGT AAA 1021
Phe Tyr Thr Ser Lys Asp Phe Leu Pro Leu Val Ala Lys Ser Ser Lys
305 310 315

GCA GGA ATG TGC GCC TGT CAC TCT CCA TTG CTG TCG ATA CGG GGA ACC 1069
Ala Gly Met Cys Ala Cys His Ser Pro Leu Leu Ser Ile Arg Gly Thr
320 325 330

GTG ATT GTA CTC GGA GCT GGA GAC ACA GCT TTC GAC TGT GCA ACA TCC 1117
Val Ile Val Leu Gly Ala Gly Asp Thr Ala Phe Asp Cys Ala Thr Ser
335 340 345 350

GCT TTA CGT TGT GGA GCC CGC CGA GTG TTC ATC GTC TTC AGA AAA GGC 1165
Ala Leu Arg Cys Gly Ala Arg Arg Val Phe Ile Val Phe Arg Lys Gly
355 360 365

TTT GTT AAT ATA AGA GCT GTC CCT GAG GAG GTG GAG CTT GCT AGA GAA 1213
Phe Val Asn Ile Arg Ala Val Pro Glu Glu Val Glu Leu Ala Arg Glu
370 375 380

GAA AAA TGT GAA TTT TTG CCT TTC TTG TCT CCA CGG AAG GTT ATA GTA 1261
Glu Lys Cys Glu Phe Leu Pro Phe Leu Ser Pro Arg Lys Val Ile Val

385 390 395

AAA GGT GGG AGA ATT GTT GCC ATG CAA TTT GTT CGG ACA GAG CAA GAT 1309
Lys Gly Gly Arg Ile Val Ala Met Gln Phe Val Arg Thr Glu Gln Asp
400 405 410

GAA ACT GGA AAA TGG AAT GAA GAT GGA GAT CAG ATA GCC TGT CTG AAA 1357
Glu Thr Gly Lys Trp Asn Glu Asp Gly Asp Gln Ile Ala Cys Leu Lys
415 420 425 430

GCC GAT GTG GTC ATC AGT GCC TTT GGC TCA GTT CTG AGT GAT CCT AAA 1405
Ala Asp Val Val Ile Ser Ala Phe Gly Ser Val Leu Ser Asp Pro Lys
435 440 445

GTA AAA GAA GCC TTG AGC CCT ATA AAA TTT AAC AGA TGG GAT CTC CCA 1453
Val Lys Glu Ala Leu Ser Pro Ile Lys Phe Asn Arg Trp Asp Leu Pro
450 455 460

GAA GTA GAT CCA GAA ACT ATG CAA ACC AGT GAG CCA TGG GTG TTT GCA 1501
Glu Val Asp Pro Glu Thr Met Gln Thr Ser Glu Pro Trp Val Phe Ala
465 470 475

GGT GGT GAT GTG GTT GGT ATA GCC AAC ACT ACA GTG GAA GCC GTG AAT 1549
Gly Gly Asp Val Val Gly Ile Ala Asn Thr Thr Val Glu Ala Val Asn
480 485 490

GAT GGA AAG CAA GCC TCT TGG TAC ATT CAC AGA TAT ATA CAG TCA CAA 1597
Asp Gly Lys Gln Ala Ser Trp Tyr Ile His Arg Tyr Ile Gln Ser Gln
495 500 505 510

TAT GGA GCT TCA GTT TCT GCT AAG CCC GAA CTC CCC CTG TTT TAT ACT 1645
Tyr Gly Ala Ser Val Ser Ala Lys Pro Glu Leu Pro Leu Phe Tyr Thr
515 520 525

CCC ATT GAT CTG GTG GAC ATT AGT GTG GAA ATG GCT GCA TTG AAG TTT 1693
Pro Ile Asp Leu Val Asp Ile Ser Val Glu Met Ala Ala Leu Lys Phe

530 535 540

ACA AAT CCT TTT GGT CTT GCC AGT GCA ACT CCA ACT ACC AGT TCG TCA 1741
Thr Asn Pro Phe Gly Leu Ala Ser Ala Thr Pro Thr Thr Ser Ser Ser
545 550 555

ATG ATT CGA AGA GCT TTT GAA GCT GGA TGG GCC TTT GCT CTG ACC AAA 1789
Met Ile Arg Arg Ala Phe Glu Ala Gly Trp Ala Phe Ala Leu Thr Lys
560 565 570

ACT TTC TCT CTT GAT AAG GAT ATA GTG ACA AAT GTT TCA CCC AGA ATC 1837
Thr Phe Ser Leu Asp Lys Asp Ile Val Thr Asn Val Ser Pro Arg Ile
575 580 585 590

ATC CGG GGG ACC ACC TCT GGC CCC ATG TAT GGC CCT GGA CAA AGC TCT 1885
Ile Arg Gly Thr Thr Ser Gly Pro Met Tyr Gly Pro Gly Gln Ser Ser
595 600 605

TTC CTG AAT ATT GAG CTC ATC AGT GAA AAA ACG GCT GCA TAT TGG TGT 1933
Phe Leu Asn Ile Glu Leu Ile Ser Glu Lys Thr Ala Ala Tyr Trp Cys
610 615 620

CAA AGT GTC ACT GAA CTA AAA GCC GAC TTT CCA GAC AAT ATT GTG ATT 1981
Gln Ser Val Thr Glu Leu Lys Ala Asp Phe Pro Asp Asn Ile Val Ile
625 630 635

GCT AGC ATC ATG TGC AGT TAC AAC AGA AAT GAC TGG ATG GAA CTC TCC 2029
Ala Ser Ile Met Cys Ser Tyr Asn Arg Asn Asp Trp Met Glu Leu Ser
640 645 650

AGA AAG GCT GAG GCT TCT GGA GCA GAC GCC CTG GAG TTA AAT TTA TCG 2077
Arg Lys Ala Glu Ala Ser Gly Ala Asp Ala Leu Glu Leu Asn Leu Ser
655 660 665 670

TGT CCG CAT GGC ATG GGA GAA AGA GGA ATG GGT CTG GCT TGT GGA CAG 2125
Cys Pro His Gly Met Gly Glu Arg Gly Met Gly Leu Ala Cys Gly Gln

675 680 685

GAT CCA GAG CTG GTG CGG AAC ATC TGT CGC TGG GTT AGG CAA GCT GTT 2173
Asp Pro Glu Leu Val Arg Asn Ile Cys Arg Trp Val Arg Gln Ala Val
690 695 700

CGG ATT CCT TTT TTT GCC AAG TTG ACC CCA AAT GTC ACT GAT ATT GTA 2221
Arg Ile Pro Phe Phe Ala Lys Leu Thr Pro Asn Val Thr Asp Ile Val
705 710 715

AGC ATA GCC AGA GCT GCA AAG GAA GGT GGG GCA AAT GGT GTT ACA GCT 2269
Ser Ile Ala Arg Ala Ala Lys Glu Gly Gly Ala Asn Gly Val Thr Ala
720 725 730

ACC AAC ACT GTC TCA GGT CTC ATG GGA TTA AAA GCT GAT GGC ACA CCC 2317
Thr Asn Thr Val Ser Gly Leu Met Gly Leu Lys Ala Asp Gly Thr Pro
735 740 745 750

TGG CCA GCA GTG GGC CGT GAG AAG CGG ACT ACA TAC GGC GGA GTG TCC 2365
Trp Pro Ala Val Gly Arg Glu Lys Arg Thr Thr Tyr Gly Gly Val Ser
755 760 765

GGC ACA GCC ATC AGA CCT ATT GCT TTG AGA GCT GTG ACC ACC ATT GCT 2413
Gly Thr Ala Ile Arg Pro Ile Ala Leu Arg Ala Val Thr Thr Ile Ala
770 775 780

CGT GCT TTG CCT GAA TTT CCC ATT TTG GCC ACT GGT GGA ATT GAC TCA 2461
Arg Ala Leu Pro Glu Phe Pro Ile Leu Ala Thr Gly Gly Ile Asp Ser
785 790 795

GCT GAA AGT GGT CTT CAG TTT CTC CAC GGT GGT GCT TCG GTG CTC CAG 2509
Ala Glu Ser Gly Leu Gln Phe Leu His Gly Gly Ala Ser Val Leu Gln
800 805 810

GTA TGC AGT GCT ATT CAA AAT CAG GAT TTC ACT ATC ATC CAA GAC TAC 2557
Val Cys Ser Ala Ile Gln Asn Gln Asp Phe Thr Ile Ile Gln Asp Tyr

815 820 825 830

TGC ACT GGC CTC AAA GCC TTG CTT TAT CTG AAA AGC ATT GAA GAA CTA 2605
Cys Thr Gly Leu Lys Ala Leu Leu Tyr Leu Lys Ser Ile Glu Glu Leu
835 840 845

CAG GAC TGG GAT GGG CAG AGT CCA GCC ACG AAG AGT CAC CAG AAA GGG 2653
Gln Asp Trp Asp Gly Gln Ser Pro Ala Thr Lys Ser His Gln Lys Gly
850 855 860

AAA CCA GTC CCT TGT ATT GCT GAA CTT GTG GGA AAG AAA CTG CCA AGC 2701
Lys Pro Val Pro Cys Ile Ala Glu Leu Val Gly Lys Lys Leu Pro Ser
865 870 875

TTT GGA CCT TAT CTT GAG AAG TGC AAG AAA ATC ATA GCA GAG GAA AAG 2749
Phe Gly Pro Tyr Leu Glu Lys Cys Lys Lys Ile Ile Ala Glu Glu Lys
880 885 890

TTG AGA CTG AAA AAA GAA AAT GTG ACC GTT CTA CCA CTT GAA AGA AAC 2797
Leu Arg Leu Lys Lys Glu Asn Val Thr Val Leu Pro Leu Glu Arg Asn
895 900 905 910

CAT TTT ATC CCA AAA AAA CCT ATT CCT TCT GTT AAG GAT GTG ATT GGA 2845
His Phe Ile Pro Lys Lys Pro Ile Pro Ser Val Lys Asp Val Ile Gly
915 920 925

AAA GCT CTG CAG TAC CTT GGA ACA TAT GGT GAA CTG AAC AAC ACA GAG 2893
Lys Ala Leu Gln Tyr Leu Gly Thr Tyr Gly Glu Leu Asn Asn Thr Glu
930 935 940

CAG GTT GTG GCT GTG ATC GAT GAA GAG ATG TGT ATC AAC TGT GGC AAA 2941
Gln Val Val Ala Val Ile Asp Glu Glu Met Cys Ile Asn Cys Gly Lys
945 950 955

TGC TAC ATG ACC TGT AAT GAC TCT GGC TAC CAG GCT ATC CAG TTT GAT 2989
Cys Tyr Met Thr Cys Asn Asp Ser Gly Tyr Gln Ala Ile Gln Phe Asp

960 965 970

CCT GAA ACC CAC CTG CCC ACC GTT ACT GAC ACT TGT ACA GGC TGT ACC 3037
Pro Glu Thr His Leu Pro Thr Val Thr Asp Thr Cys Thr Gly Cys Thr
975 980 985 990

CTG TGT CTC TCC GTC TGC CCT ATT ATC GAC TGC ATC AAA ATG GTT TCC 3085
Leu Cys Leu Ser Val Cys Pro Ile Ile Asp Cys Ile Lys Met Val Ser
995 1000 1005

AGG ACA ACA CCT TAT GAA CCA AAG AGA GGC TTG CCC TTG GCT GTG AAT 3133
Arg Thr Thr Pro Tyr Glu Pro Lys Arg Gly Leu Pro Leu Ala Val Asn
1010 1015 1020

CCT GTG TCT TAAGGTGATT TGTGAAACAG TTGCAGTGAA CTTCGAGGTC 3182
Pro Val Ser
1025

ACCTACTTAT GCTGATCTTT TCAATAGTGA TCATTATGCT CAGCTTTTTC TAAATTCAAA 3242

CATATAATTT CTAAATTTAA AAAAAGATTA TTTCTAAAGA AATTTCTAAA TTTTAAAAAT 3302

GTCTGCTTTC AGTGATCATT CAATTAATGG TCATAAAATA GAATAATTCT TTTCTGAGCA 3362

GAATTGTTCA ATATAACTAT GGAGCAGTTA ATTGGATGTT CACCATCAGT TGTCCATTAT 3422

GAAAAAATTA ACTTTTTTGT AGCAATTAAT GCTACACTTT TCAAATTGCC CTATGCCGAG 3482

TTCTGTCTTT GATTTCTAAT TGTAAGGGAA ATTAAGTATT TTAGAACAAA GTACAATTTA 3542

ACTTACAACA AATGTTTCCA AGGAAACATT TTATAATTAA AAATTACATT TAATTTTAAC 3602

TCTGTTTCTA AGCAAAAGTA ATTAGCTCCA TAAAGCTCAG ATGAAGTCAA ATAATTATTT 3662

ACTGTGGTAG CAAAAGAAAG CCAATGAGGG TTTGGGAAAC TTTCCTTAAG GTCTCTTCAC 3722

TGAAATAACT GGATACTGAA GGCGAGAGTG TTCAGTAACC ATTTGTATCA AGCTATGCTA 3782

TTCACCGCTC AGGCCTGAGA TGTGTGTCCA AATGCTACCA ATGAATCAAC ATGACATTCC 3842

TGTTTAAATA TTTAAACTAT GTTCCTAACA AAGTAAGACA TTAGGATGGA ACTCTGGTTA 3902

AAGCCACTCT TTTGCTGTGC ACAGATCTGT TCTATCTGCT TCTAATATAG TCACCTTCGT 3962

GATCCTAGCA ATTAATGTTT GAACACAGCA CAGATTATAC AGAAGTGGGG TCATGTGCTT 4022

CTTTATTCAA GAATGAGAAA TCCAGTATGG GTAATATATA TTATATGGGT GATACCACTT 4082

TACCAACTCT TTATTTTAGT GTCCATGTTG AATTTCGAAA GTGATTAAAA AAGAAATGGT 4142

ATTTTCTGTT ACTGCCAAAT AATATTTTTA TATTCCTCGA TTTTTAAAAT CAGCAAATAG 4202

CATCTTATAA ACTTGTTTAT CTCTTCTTTG TGGCATATTT TAATATGAAT CCATAAGTAG 4262

TAAATCTTCA TGTAATCATC CATAGCACCT TTCTATGACA AATGCAAGAT CAAGAGAAAA 4322

ATAAATGTTT GATTATGCAC TTTTAGAAAT GCACATTTAC CACAAAATCT GTATGATCAA 4382

ATAATATTAA ATAAAATTTT ATAAAGCATT TT 4414

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1025 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Pro Val Leu Ser Lys Asp Val Ala Asp Ile Glu Ser Ile Leu
1 5 10 15

Ala Leu Asn Pro Arg Thr Gln Ser Arg Ala Thr Leu Arg Ser Thr Leu
20 25 30

Ala Lys Lys Leu Asp Lys Lys His Trp Lys Arg Asn Pro Asp Lys Asn
35 40 45

Cys Phe Asn Cys Glu Lys Leu Glu Asn Asn Phe Asp Asp Ile Lys His
50 55 60

Thr Thr Leu Gly Glu Arg Gly Ala Leu Arg Glu Ala Met Arg Cys Leu
65 70 75 80

Lys Cys Ala Asp Ala Pro Cys Gln Lys Ser Cys Pro Thr Asn Leu Asp
85 90 95

Ile Lys Ser Phe Ile Thr Ser Ile Ser Asn Lys Asn Tyr Tyr Gly Ala
100 105 110

Ala Lys Met Ile Phe Ser Asp Asn Pro Leu Gly Leu Thr Cys Gly Met
115 120 125

Val Cys Pro Thr Ser Asp Leu Cys Val Gly Gly Cys Asn Leu Tyr Ala
130 135 140

Thr Glu Glu Gly Pro Ile Asn Ile Gly Gly Leu Gln Gln Tyr Ala Thr
145 150 155 160

Glu Val Phe Lys Ala Met Asn Ile Pro Gln Ile Arg Asn Pro Ser Leu
165 170 175

Pro Pro Pro Glu Lys Met Pro Glu Ala Tyr Ser Ala Lys Ile Ala Leu
180 185 190

Leu Gly Ala Gly Pro Ala Ser Ile Ser Cys Ala Ser Phe Leu Ala Arg
195 200 205

Leu Gly Tyr Asn Asp Ile Thr Ile Phe Glu Lys Gln Glu Tyr Val Gly
210 215 220

Gly Ile Ser Thr Ser Glu Ile Pro Gln Phe Arg Leu Pro Tyr Asp Val
225 230 235 240

Val Asn Phe Glu Ile Glu Leu Met Lys Asp Leu Gly Val Lys Ile Ile
245 250 255

Cys Gly Lys Ser Leu Ser Val Asn Asp Ile Thr Leu Ser Thr Leu Lys
260 265 270

Glu Glu Gly Tyr Lys Ala Ala Phe Ile Gly Ile Gly Leu Pro Glu Pro
275 280 285

Lys Lys Asp His Ile Phe Gln Gly Leu Thr Gln Asp Gln Gly Phe Tyr
290 295 300

Thr Ser Lys Asp Phe Leu Pro Leu Val Ala Lys Ser Ser Lys Ala Gly
305 310 315 320

Met Cys Ala Cys His Ser Pro Leu Leu Ser Ile Arg Gly Thr Val Ile
325 330 335

Val Leu Gly Ala Gly Asp Thr Ala Phe Asp Cys Ala Thr Ser Ala Leu
340 345 350

Arg Cys Gly Ala Arg Arg Val Phe Ile Val Phe Arg Lys Gly Phe Val
355 360 365

Asn Ile Arg Ala Val Pro Glu Glu Val Glu Leu Ala Arg Glu Glu Lys
370 375 380

Cys Glu Phe Leu Pro Phe Leu Ser Pro Arg Lys Val Ile Val Lys Gly
385 390 395 400

Gly Arg Ile Val Ala Met Gln Phe Val Arg Thr Glu Gln Asp Glu Thr
405 410 415

Gly Lys Trp Asn Glu Asp Gly Asp Gln Ile Ala Cys Leu Lys Ala Asp
420 425 430

Val Val Ile Ser Ala Phe Gly Ser Val Leu Ser Asp Pro Lys Val Lys
435 440 445

Glu Ala Leu Ser Pro Ile Lys Phe Asn Arg Trp Asp Leu Pro Glu Val
450 455 460

Asp Pro Glu Thr Met Gln Thr Ser Glu Pro Trp Val Phe Ala Gly Gly
465 470 475 480

Asp Val Val Gly Ile Ala Asn Thr Thr Val Glu Ala Val Asn Asp Gly
485 490 495

Lys Gln Ala Ser Trp Tyr Ile His Arg Tyr Ile Gln Ser Gln Tyr Gly
500 505 510

Ala Ser Val Ser Ala Lys Pro Glu Leu Pro Leu Phe Tyr Thr Pro Ile
515 520 525

Asp Leu Val Asp Ile Ser Val Glu Met Ala Ala Leu Lys Phe Thr Asn
530 535 540

Pro Phe Gly Leu Ala Ser Ala Thr Pro Thr Thr Ser Ser Ser Met Ile
545 550 555 560

Arg Arg Ala Phe Glu Ala Gly Trp Ala Phe Ala Leu Thr Lys Thr Phe
565 570 575

Ser Leu Asp Lys Asp Ile Val Thr Asn Val Ser Pro Arg Ile Ile Arg
580 585 590

Gly Thr Thr Ser Gly Pro Met Tyr Gly Pro Gly Gln Ser Ser Phe Leu
595 600 605

Asn Ile Glu Leu Ile Ser Glu Lys Thr Ala Ala Tyr Trp Cys Gln Ser
610 615 620

Val Thr Glu Leu Lys Ala Asp Phe Pro Asp Asn Ile Val Ile Ala Ser
625 630 635 640

Ile Met Cys Ser Tyr Asn Arg Asn Asp Trp Met Glu Leu Ser Arg Lys
645 650 655

Ala Glu Ala Ser Gly Ala Asp Ala Leu Glu Leu Asn Leu Ser Cys Pro
660 665 670

His Gly Met Gly Glu Arg Gly Met Gly Leu Ala Cys Gly Gln Asp Pro
675 680 685

Glu Leu Val Arg Asn Ile Cys Arg Trp Val Arg Gln Ala Val Arg Ile
690 695 700

Pro Phe Phe Ala Lys Leu Thr Pro Asn Val Thr Asp Ile Val Ser Ile
705 710 715 720

Ala Arg Ala Ala Lys Glu Gly Gly Ala Asn Gly Val Thr Ala Thr Asn
725 730 735

Thr Val Ser Gly Leu Met Gly Leu Lys Ala Asp Gly Thr Pro Trp Pro
740 745 750

Ala Val Gly Arg Glu Lys Arg Thr Thr Tyr Gly Gly Val Ser Gly Thr
755 760 765

Ala Ile Arg Pro Ile Ala Leu Arg Ala Val Thr Thr Ile Ala Arg Ala
770 775 780

Leu Pro Glu Phe Pro Ile Leu Ala Thr Gly Gly Ile Asp Ser Ala Glu
785 790 795 800

Ser Gly Leu Gln Phe Leu His Gly Gly Ala Ser Val Leu Gln Val Cys
805 810 815

Ser Ala Ile Gln Asn Gln Asp Phe Thr Ile Ile Gln Asp Tyr Cys Thr
820 825 830

Gly Leu Lys Ala Leu Leu Tyr Leu Lys Ser Ile Glu Glu Leu Gln Asp
835 840 845

Trp Asp Gly Gln Ser Pro Ala Thr Lys Ser His Gln Lys Gly Lys Pro
850 855 860

Val Pro Cys Ile Ala Glu Leu Val Gly Lys Lys Leu Pro Ser Phe Gly
865 870 875 880

Pro Tyr Leu Glu Lys Cys Lys Lys Ile Ile Ala Glu Glu Lys Leu Arg
885 890 895

Leu Lys Lys Glu Asn Val Thr Val Leu Pro Leu Glu Arg Asn His Phe
900 905 910

Ile Pro Lys Lys Pro Ile Pro Ser Val Lys Asp Val Ile Gly Lys Ala
915 920 925

Leu Gln Tyr Leu Gly Thr Tyr Gly Glu Leu Asn Asn Thr Glu Gln Val
930 935 940

Val Ala Val Ile Asp Glu Glu Met Cys Ile Asn Cys Gly Lys Cys Tyr
945 950 955 960

Met Thr Cys Asn Asp Ser Gly Tyr Gln Ala Ile Gln Phe Asp Pro Glu
965 970 975

Thr His Leu Pro Thr Val Thr Asp Thr Cys Thr Gly Cys Thr Leu Cys
980 985 990

Leu Ser Val Cys Pro Ile Ile Asp Cys Ile Lys Met Val Ser Arg Thr
995 1000 1005

Thr Pro Tyr Glu Pro Lys Arg Gly Leu Pro Leu Ala Val Asn Pro Val
1010 1015 1020

Ser
1025

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4368 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 49..3123

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ACGCAAGGAG GGTTTGTCAC TGGCAGACTC GAGACTGTAG GCACTGCC ATG GCC CCT 57
Met Ala Pro

GTG CTC AGT AAG GAC TCG GCG GAC ATC GAG AGT ATC CTG GCT TTA AAT 105
Val Leu Ser Lys Asp Ser Ala Asp Ile Glu Ser Ile Leu Ala Leu Asn
1030 1035 1040

CCT CGA ACA CAA ACT CAT GCA ACT CTG TGT TCC ACT TCG GCC AAG AAA 153
Pro Arg Thr Gln Thr His Ala Thr Leu Cys Ser Thr Ser Ala Lys Lys
1045 1050 1055 1060

TTA GAC AAG AAA CAT TGG AAA AGA AAT CCT GAT AAG AAC TGC TTT AAT 201
Leu Asp Lys Lys His Trp Lys Arg Asn Pro Asp Lys Asn Cys Phe Asn
1065 1070 1075

TGT GAG AAG CTG GAG AAT AAT TTT GAT GAC ATC AAG CAC ACG ACT CTT 249
Cys Glu Lys Leu Glu Asn Asn Phe Asp Asp Ile Lys His Thr Thr Leu
1080 1085 1090

GGT GAG CGA GGA GCT CTC CGA GAA GCA ATG AGA TGC CTG AAA TGT GCA 297
Gly Glu Arg Gly Ala Leu Arg Glu Ala Met Arg Cys Leu Lys Cys Ala
1095 1100 1105

GAT GCC CCG TGT CAG AAG AGC TGT CCA ACT AAT CTT GAT ATT AAA TCA 345
Asp Ala Pro Cys Gln Lys Ser Cys Pro Thr Asn Leu Asp Ile Lys Ser
1110 1115 1120

TTC ATC ACA AGT ATT GCA AAC AAG AAC TAT TAT GGA GCT GCT AAG ATG 393
Phe Ile Thr Ser Ile Ala Asn Lys Asn Tyr Tyr Gly Ala Ala Lys Met
1125 1130 1135 1140

ATA TTT TCT GAC AAC CCA CTT GGT CTG ACT TGT GGA ATG GTA TGT CCA 441
Ile Phe Ser Asp Asn Pro Leu Gly Leu Thr Cys Gly Met Val Cys Pro

1145 1150 1155

ACC TCT GAT CTT TGT GTA GGT GGA TGC AAT TTA TAT GCC ACT GAA GAG 489
Thr Ser Asp Leu Cys Val Gly Gly Cys Asn Leu Tyr Ala Thr Glu Glu
1160 1165 1170

GGA CCC ATT AAT ATT GGT GGA TTG CAG CAA TTT GCT ACT GAG GTA TTC 537
Gly Pro Ile Asn Ile Gly Gly Leu Gln Gln Phe Ala Thr Glu Val Phe
1175 1180 1185

AAA GCA ATG AGT ATC CCA CAG ATC AGA AAT CCT TCG CTG CCT CCC CCA 585
Lys Ala Met Ser Ile Pro Gln Ile Arg Asn Pro Ser Leu Pro Pro Pro
1190 1195 1200

GAA AAA ATG TCT GAA GCC TAT TCT GCA AAG ATT GCT CTT TTT GGT GCT 633
Glu Lys Met Ser Glu Ala Tyr Ser Ala Lys Ile Ala Leu Phe Gly Ala
1205 1210 1215 1220

GGG CCT GCA AGT ATA AGT TGT GCT TCC TTT TTG GCT CGA TTG GGG TAC 681
Gly Pro Ala Ser Ile Ser Cys Ala Ser Phe Leu Ala Arg Leu Gly Tyr
1225 1230 1235

TCT GAC ATC ACT ATA TTT GAA AAA CAA GAA TAT GTT GGT GGT TTA AGT 729
Ser Asp Ile Thr Ile Phe Glu Lys Gln Glu Tyr Val Gly Gly Leu Ser
1240 1245 1250

ACT TCT GAA ATT CCT CAG TTC CGG CTG CCG TAT GAT GTA GTG AAT TTT 777
Thr Ser Glu Ile Pro Gln Phe Arg Leu Pro Tyr Asp Val Val Asn Phe
1255 1260 1265

GAG ATT GAG CTA ATG AAG GAC CTT GGT GTA AAG ATA ATT TGC GGT AAA 825
Glu Ile Glu Leu Met Lys Asp Leu Gly Val Lys Ile Ile Cys Gly Lys
1270 1275 1280

AGC CTT TCA GTG AAT GAA ATG ACT CTT AGC ACT TTG AAA GAA AAA GGC 873
Ser Leu Ser Val Asn Glu Met Thr Leu Ser Thr Leu Lys Glu Lys Gly

1285 1290 1295 1300

TAC AAA GCT GCT TTC ATT GGA ATA GGT TTG CCA GAA CCC AAT AAA GAT 921
Tyr Lys Ala Ala Phe Ile Gly Ile Gly Leu Pro Glu Pro Asn Lys Asp
1305 1310 1315

GCC ATC TTC CAA GGC CTG ACG CAG GAC CAG GGG TTT TAT ACA TCC AAA 969
Ala Ile Phe Gln Gly Leu Thr Gln Asp Gln Gly Phe Tyr Thr Ser Lys
1320 1325 1330

GAC TTT TTG CCA CTT GTA GCC AAA GGC AGT AAA GCA GGA ATG TGC GCC 1017
Asp Phe Leu Pro Leu Val Ala Lys Gly Ser Lys Ala Gly Met Cys Ala
1335 1340 1345

TGT CAC TCT CCA TTG CCA TCG ATA CGG GGA GTC GTG ATT GTA CTT GGA 1065
Cys His Ser Pro Leu Pro Ser Ile Arg Gly Val Val Ile Val Leu Gly
1350 1355 1360

GCT GGA GAC ACT GCC TTT GAC TGT GCA ACA TCT GCT CTA CGT TGT GGA 1113
Ala Gly Asp Thr Ala Phe Asp Cys Ala Thr Ser Ala Leu Arg Cys Gly
1365 1370 1375 1380

GCT CGC CGT GTG TTC ATC GTC TTC AGA AAA GGC TTT GTT AAT ATA AGA 1161
Ala Arg Arg Val Phe Ile Val Phe Arg Lys Gly Phe Val Asn Ile Arg
1385 1390 1395

GCT GTC CCT GAG GAG ATG GAA CTT GCT AAG GAA GAA AAG TGT GAA TTT 1209
Ala Val Pro Glu Glu Met Glu Leu Ala Lys Glu Glu Lys Cys Glu Phe
1400 1405 1410

CTG CCA TTC CTG TCC CCA CGG AAG GTT ATA GTA AAA GGT GGG AGA ATT 1257
Leu Pro Phe Leu Ser Pro Arg Lys Val Ile Val Lys Gly Gly Arg Ile
1415 1420 1425

GTT GCT ATG CAG TTT GTT CGG ACA GAG CAA GAT GAA ACT GGA AAA TGG 1305
Val Ala Met Gln Phe Val Arg Thr Glu Gln Asp Glu Thr Gly Lys Trp

1430 1435 1440

AAT GAA GAT GAA GAT CAG ATG GTC CAT CTG AAA GCC GAT GTG GTC ATC 1353
Asn Glu Asp Glu Asp Gln Met Val His Leu Lys Ala Asp Val Val Ile
1445 1450 1455 1460

AGT GCC TTT GGT TCA GTT CTG AGT GAT CCT AAA GTA AAA GAA GCC TTG 1401
Ser Ala Phe Gly Ser Val Leu Ser Asp Pro Lys Val Lys Glu Ala Leu
1465 1470 1475

AGC CCT ATA AAA TTT AAC AGA TGG GGT CTC CCA GAA GTA GAT CCA GAA 1449
Ser Pro Ile Lys Phe Asn Arg Trp Gly Leu Pro Glu Val Asp Pro Glu
1480 1485 1490

ACT ATG CAA ACT AGT GAA GCA TGG GTA TTT GCA GGT GGT GAT GTC GTT 1497
Thr Met Gln Thr Ser Glu Ala Trp Val Phe Ala Gly Gly Asp Val Val
1495 1500 1505

GGT TTG GCT AAC ACT ACA GTG GAA TCG GTG AAT GAT GGA AAG CAA GCT 1545
Gly Leu Ala Asn Thr Thr Val Glu Ser Val Asn Asp Gly Lys Gln Ala
1510 1515 1520

TCT TGG TAC ATT CAC AAA TAC GTA CAG TCA CAA TAT GGA GCT TCC GTT 1593
Ser Trp Tyr Ile His Lys Tyr Val Gln Ser Gln Tyr Gly Ala Ser Val
1525 1530 1535 1540

TCT GCC AAG CCT GAA CTA CCC CTC TTT TAC ACT CCT ATT GAT CTG GTG 1641
Ser Ala Lys Pro Glu Leu Pro Leu Phe Tyr Thr Pro Ile Asp Leu Val
1545 1550 1555

GAC ATT AGT GTA GAA ATG GCC GGA TTG AAG TTT ATA AAT CCT TTT GGT 1689
Asp Ile Ser Val Glu Met Ala Gly Leu Lys Phe Ile Asn Pro Phe Gly
1560 1565 1570

CTT GCT AGC GCA ACT CCA GCC ACC AGC ACA TCA ATG ATT CGA AGA GCT 1737
Leu Ala Ser Ala Thr Pro Ala Thr Ser Thr Ser Met Ile Arg Arg Ala

1575 1580 1585

TTT GAA GCT GGA TGG GGT TTT GCC CTC ACC AAA ACT TTC TCT CTT GAT 1785
Phe Glu Ala Gly Trp Gly Phe Ala Leu Thr Lys Thr Phe Ser Leu Asp
1590 1595 1600

AAG GAC ATT GTG ACA AAT GTT TCC CCC AGA ATC ATC CGG GGA ACC ACC 1833
Lys Asp Ile Val Thr Asn Val Ser Pro Arg Ile Ile Arg Gly Thr Thr
1605 1610 1615 1620

TCT GGC CCC ATG TAT GGC CCT GGA CAA AGC TCC TTT CTG AAT ATT GAG 1881
Ser Gly Pro Met Tyr Gly Pro Gly Gln Ser Ser Phe Leu Asn Ile Glu
1625 1630 1635

CTC ATC AGT GAG AAA ACG GCT GCA TAT TGG TGT CAA AGT GTC ACT GAA 1929
Leu Ile Ser Glu Lys Thr Ala Ala Tyr Trp Cys Gln Ser Val Thr Glu
1640 1645 1650

CTA AAG GCT GAC TTT CCA GAC AAC ATT GTG ATT GCT AGC ATT ATG TGC 1977
Leu Lys Ala Asp Phe Pro Asp Asn Ile Val Ile Ala Ser Ile Met Cys
1655 1660 1665

AGT TAC AAT AAA AAT GAC TGG ACG GAA CTT GCC AAG AAG TCT GAG GAT 2025
Ser Tyr Asn Lys Asn Asp Trp Thr Glu Leu Ala Lys Lys Ser Glu Asp
1670 1675 1680

TCT GGA GCA GAT GCC CTG GAG TTA AAT TTA TCA TGT CCA CAT GGC ATG 2073
Ser Gly Ala Asp Ala Leu Glu Leu Asn Leu Ser Cys Pro His Gly Met
1685 1690 1695 1700

GGA GAA AGA GGA ATG GGC CTG GCC TGT GGG CAG GAT CCA GAG CTG GTG 2121
Gly Glu Arg Gly Met Gly Leu Ala Cys Gly Gln Asp Pro Glu Leu Val
1705 1710 1715

CGG AAC ATC TGC CGC TGG GTT AGG CAA GCT GTT CAG ATT CCT TTT TTT 2169
Arg Asn Ile Cys Arg Trp Val Arg Gln Ala Val Gln Ile Pro Phe Phe

1720 1725 1730

GCC AAG CTG ACC CCA AAT GTC ACT GAT ATT GTG AGC ATC GCA AGA GCT 2217
Ala Lys Leu Thr Pro Asn Val Thr Asp Ile Val Ser Ile Ala Arg Ala
1735 1740 1745

GCA AAG GAA GGT GGT GCC AAT GGC GTT ACA GCC ACC AAC ACT GTC TCA 2265
Ala Lys Glu Gly Gly Ala Asn Gly Val Thr Ala Thr Asn Thr Val Ser
1750 1755 1760

GGT CTG ATG GGA TTA AAA TCT GAT GGC ACA CCT TGG CCA GCA GTG GGG 2313
Gly Leu Met Gly Leu Lys Ser Asp Gly Thr Pro Trp Pro Ala Val Gly
1765 1770 1775 1780

ATT GCA AAG CGA ACT ACA TAT GGA GGA GTG TCT GGG ACA GCA ATC AGA 2361
Ile Ala Lys Arg Thr Thr Tyr Gly Gly Val Ser Gly Thr Ala Ile Arg
1785 1790 1795

CCT ATT GCT TTG AGA GCT GTG ACC TCC ATT GCT CGT GCT CTG CCT GGA 2409
Pro Ile Ala Leu Arg Ala Val Thr Ser Ile Ala Arg Ala Leu Pro Gly
1800 1805 1810

TTT CCC ATT TTG GCT ACT GGT GGA ATT GAC TCT GCT GAA AGT GGT CTT 2457
Phe Pro Ile Leu Ala Thr Gly Gly Ile Asp Ser Ala Glu Ser Gly Leu
1815 1820 1825

CAG TTT CTC CAT AGT GGT GCT TCC GTC CTC CAG GTA TGC AGT GCC ATT 2505
Gln Phe Leu His Ser Gly Ala Ser Val Leu Gln Val Cys Ser Ala Ile
1830 1835 1840

CAG AAT CAG GAT TTC ACT GTG ATC GAA GAC TAC TGC ACT GGC CTC AAA 2553
Gln Asn Gln Asp Phe Thr Val Ile Glu Asp Tyr Cys Thr Gly Leu Lys
1845 1850 1855 1860

GCC CTG CTT TAT CTG AAA AGC ATT GAA GAA CTA CAA GAC TGG GAT GGA 2601
Ala Leu Leu Tyr Leu Lys Ser Ile Glu Glu Leu Gln Asp Trp Asp Gly

1865 1870 1875

CAG AGT CCA GCT ACT GTG AGT CAC CAG AAA GGG AAA CCA GTT CCA CGT 2649
Gln Ser Pro Ala Thr Val Ser His Gln Lys Gly Lys Pro Val Pro Arg
1880 1885 1890

ATA GCT GAA CTC ATG GAC AAG AAA CTG CCA AGT TTT GGA CCT TAT CTG 2697
Ile Ala Glu Leu Met Asp Lys Lys Leu Pro Ser Phe Gly Pro Tyr Leu
1895 1900 1905

GAA CAG CGC AAG AAA ATC ATA GCA GAA AAC AAG ATT AGA CTG AAA GAA 2745
Glu Gln Arg Lys Lys Ile Ile Ala Glu Asn Lys Ile Arg Leu Lys Glu
1910 1915 1920

CAA AAT GTA GCT TTT TCA CCA CTT AAG AGA AAC TGT TTT ATC CCC AAA 2793
Gln Asn Val Ala Phe Ser Pro Leu Lys Arg Asn Cys Phe Ile Pro Lys
1925 1930 1935 1940

AGG CCT ATT CCT ACC ATC AAG GAT GTA ATA GGA AAA GCA CTG CAG TAC 2841
Arg Pro Ile Pro Thr Ile Lys Asp Val Ile Gly Lys Ala Leu Gln Tyr
1945 1950 1955

CTT GGA ACA TTT GGT GAA TTG AGC AAC GTA GAG CAA GTT GTG GCT ATG 2889
Leu Gly Thr Phe Gly Glu Leu Ser Asn Val Glu Gln Val Val Ala Met
1960 1965 1970

ATT GAT GAA GAA ATG TGT ATC AAC TGT GGT AAA TGC TAC ATG ACC TGT 2937
Ile Asp Glu Glu Met Cys Ile Asn Cys Gly Lys Cys Tyr Met Thr Cys
1975 1980 1985

AAT GAT TCT GGC TAC CAG GCT ATA CAG TTT GAT CCA GAA ACC CAC CTG 2985
Asn Asp Ser Gly Tyr Gln Ala Ile Gln Phe Asp Pro Glu Thr His Leu
1990 1995 2000

CCC ACC ATA ACC GAC ACT TGT ACA GGC TGT ACT CTG TGT CTC AGT GTT 3033
Pro Thr Ile Thr Asp Thr Cys Thr Gly Cys Thr Leu Cys Leu Ser Val

2005 2010 2015 2020

TGC CCT ATT GTC GAC TGC ATC AAA ATG GTT TCC AGG ACA ACA CCT TAT 3081
Cys Pro Ile Val Asp Cys Ile Lys Met Val Ser Arg Thr Thr Pro Tyr
2025 2030 2035

GAA CCA AAG AGA GGC GTA CCC TTA TCT GTG AAT CCG GTG TGT 3123
Glu Pro Lys Arg Gly Val Pro Leu Ser Val Asn Pro Val Cys
2040 2045 2050

TAAGGTGATT TGTGAAACAG TTGCTGTGAA CTTTCATGTC ACCTACATAT GCTGATCTTT 3183

TAAAATCATG ATCCTTGTGT TCAGCTCTTT CCAAATTAAA ACAAATATAC ATTTTCTAAA 3243

TAAAAATATG TAATTTCAAA ATACATTTGT AAGTGTAAAA AATGTCTCAT GTCAATGACC 3303

ATTCAATTAG TGGTCATAAA ATAGAATAAT TCTTTTCTGA GGATAGTAGT TAAATAACTG 3363

TGTGGCAGTT AATTGGATGT TCACTGCCAG TTGTCTTATG TGAAAAATTA ACTTTTTTGT 3423

GGCAATTAGT GTGACAGTTT CCAAATTGCC CTATGCTGTG CTCCATATTT GATTTCTAAT 3483

TGTAAGTGAA ATTAAGCATT TTGAAACAAA GTACTCTTTA ACATACAAGA AAATGTATCC 3543

AAGGAAACAT TTTATCATTA AAAATTACCT TTAATTTTAA TGCTGTTTCT AAGAAAATGT 3603

AGTTAGCTCC ATAAAGTACA AATGAAGAAA GTCAAAAAAT TATTTGCTAT GGCAGGATAA 3663

GAAAGCCTAA AATTGAGTTT GTAGAACTTT ATTAAGTAAA ATCCCCTTCG CTGAAATTGC 3723

TTATTTTTGG TGTTGGATAG AGGATAGGGA GAATATTTAC TAACTAAATA CCATTCACTA 3783

CTCATGCGTG AGATGGGTGT ACAAACTCAT CCTCTTTTAA TGGCATTTCT CTTTAAACTA 3843

TGTTCCTAAC AAAATGAGAT GATAGGATAG ATCCTGGTTA CCACTCTTTT GCTGTGCACA 3903

TACGGGCTCT GACTGGTTTT AATAGTCACC TTCATGATTA TAGCAACTAA TGTTTGAACA 3963

AAGCTCAAAG TATGCAATGC TTCATTATTC AAGAATGAAA AATATAATGT TGATAATATA 4023

TATTAAGTGT GCCAAATCAG TTTGACTACT CTCTGTTTTA GTGTTTATGT TTAAAAGAAA 4083

TATATTTTTT GTTATTATTA GATAATATTT TTGTATTTCT CTATTTTCAT AATCAGTAAA 4143

TAGTGTCATA TAAACTCATT TATCTCCTCT TCATGGCATC TTCAATATGA ATCTATAAGT 4203

AGTAAATCAG AAAGTAACAA TCTATGGCTT ATTTCTATGA CAAATTCAAG AGCTAGAAAA 4263

ATAAAATGTT TCATTATGCA CTTTTAGAAA TGCATATTTG CCACAAAACC TGTATTACTG 4323

AATAATATCA AATAAAATAT CATAAAGCAT TTTAAAAAAA AAAAA 4368

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1025 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ala Pro Val Leu Ser Lys Asp Ser Ala Asp Ile Glu Ser Ile Leu
1 5 10 15

Ala Leu Asn Pro Arg Thr Gln Thr His Ala Thr Leu Cys Ser Thr Ser
20 25 30

Ala Lys Lys Leu Asp Lys Lys His Trp Lys Arg Asn Pro Asp Lys Asn
35 40 45

Cys Phe Asn Cys Glu Lys Leu Glu Asn Asn Phe Asp Asp Ile Lys His
50 55 60

Thr Thr Leu Gly Glu Arg Gly Ala Leu Arg Glu Ala Met Arg Cys Leu
65 70 75 80

Lys Cys Ala Asp Ala Pro Cys Gln Lys Ser Cys Pro Thr Asn Leu Asp
85 90 95

Ile Lys Ser Phe Ile Thr Ser Ile Ala Asn Lys Asn Tyr Tyr Gly Ala
100 105 110

Ala Lys Met Ile Phe Ser Asp Asn Pro Leu Gly Leu Thr Cys Gly Met
115 120 125

Val Cys Pro Thr Ser Asp Leu Cys Val Gly Gly Cys Asn Leu Tyr Ala
130 135 140

Thr Glu Glu Gly Pro Ile Asn Ile Gly Gly Leu Gln Gln Phe Ala Thr
145 150 155 160

Glu Val Phe Lys Ala Met Ser Ile Pro Gln Ile Arg Asn Pro Ser Leu
165 170 175

Pro Pro Pro Glu Lys Met Ser Glu Ala Tyr Ser Ala Lys Ile Ala Leu
180 185 190

Phe Gly Ala Gly Pro Ala Ser Ile Ser Cys Ala Ser Phe Leu Ala Arg
195 200 205

Leu Gly Tyr Ser Asp Ile Thr Ile Phe Glu Lys Gln Glu Tyr Val Gly
210 215 220

Gly Leu Ser Thr Ser Glu Ile Pro Gln Phe Arg Leu Pro Tyr Asp Val
225 230 235 240

Val Asn Phe Glu Ile Glu Leu Met Lys Asp Leu Gly Val Lys Ile Ile
245 250 255

Cys Gly Lys Ser Leu Ser Val Asn Glu Met Thr Leu Ser Thr Leu Lys
260 265 270

Glu Lys Gly Tyr Lys Ala Ala Phe Ile Gly Ile Gly Leu Pro Glu Pro
275 280 285

Asn Lys Asp Ala Ile Phe Gln Gly Leu Thr Gln Asp Gln Gly Phe Tyr
290 295 300

Thr Ser Lys Asp Phe Leu Pro Leu Val Ala Lys Gly Ser Lys Ala Gly
305 310 315 320

Met Cys Ala Cys His Ser Pro Leu Pro Ser Ile Arg Gly Val Val Ile
325 330 335

Val Leu Gly Ala Gly Asp Thr Ala Phe Asp Cys Ala Thr Ser Ala Leu
340 345 350

Arg Cys Gly Ala Arg Arg Val Phe Ile Val Phe Arg Lys Gly Phe Val
355 360 365

Asn Ile Arg Ala Val Pro Glu Glu Met Glu Leu Ala Lys Glu Glu Lys
370 375 380

Cys Glu Phe Leu Pro Phe Leu Ser Pro Arg Lys Val Ile Val Lys Gly
385 390 395 400

Gly Arg Ile Val Ala Met Gln Phe Val Arg Thr Glu Gln Asp Glu Thr
405 410 415

Gly Lys Trp Asn Glu Asp Glu Asp Gln Met Val His Leu Lys Ala Asp
420 425 430

Val Val Ile Ser Ala Phe Gly Ser Val Leu Ser Asp Pro Lys Val Lys
435 440 445

Glu Ala Leu Ser Pro Ile Lys Phe Asn Arg Trp Gly Leu Pro Glu Val
450 455 460

Asp Pro Glu Thr Met Gln Thr Ser Glu Ala Trp Val Phe Ala Gly Gly
465 470 475 480

Asp Val Val Gly Leu Ala Asn Thr Thr Val Glu Ser Val Asn Asp Gly
485 490 495

Lys Gln Ala Ser Trp Tyr Ile His Lys Tyr Val Gln Ser Gln Tyr Gly
500 505 510

Ala Ser Val Ser Ala Lys Pro Glu Leu Pro Leu Phe Tyr Thr Pro Ile
515 520 525

Asp Leu Val Asp Ile Ser Val Glu Met Ala Gly Leu Lys Phe Ile Asn
530 535 540

Pro Phe Gly Leu Ala Ser Ala Thr Pro Ala Thr Ser Thr Ser Met Ile
545 550 555 560

Arg Arg Ala Phe Glu Ala Gly Trp Gly Phe Ala Leu Thr Lys Thr Phe
565 570 575

Ser Leu Asp Lys Asp Ile Val Thr Asn Val Ser Pro Arg Ile Ile Arg
580 585 590

Gly Thr Thr Ser Gly Pro Met Tyr Gly Pro Gly Gln Ser Ser Phe Leu
595 600 605

Asn Ile Glu Leu Ile Ser Glu Lys Thr Ala Ala Tyr Trp Cys Gln Ser
610 615 620

Val Thr Glu Leu Lys Ala Asp Phe Pro Asp Asn Ile Val Ile Ala Ser
625 630 635 640

Ile Met Cys Ser Tyr Asn Lys Asn Asp Trp Thr Glu Leu Ala Lys Lys
645 650 655

Ser Glu Asp Ser Gly Ala Asp Ala Leu Glu Leu Asn Leu Ser Cys Pro
660 665 670

His Gly Met Gly Glu Arg Gly Met Gly Leu Ala Cys Gly Gln Asp Pro
675 680 685

Glu Leu Val Arg Asn Ile Cys Arg Trp Val Arg Gln Ala Val Gln Ile
690 695 700

Pro Phe Phe Ala Lys Leu Thr Pro Asn Val Thr Asp Ile Val Ser Ile
705 710 715 720

Ala Arg Ala Ala Lys Glu Gly Gly Ala Asn Gly Val Thr Ala Thr Asn
725 730 735

Thr Val Ser Gly Leu Met Gly Leu Lys Ser Asp Gly Thr Pro Trp Pro
740 745 750

Ala Val Gly Ile Ala Lys Arg Thr Thr Tyr Gly Gly Val Ser Gly Thr
755 760 765

Ala Ile Arg Pro Ile Ala Leu Arg Ala Val Thr Ser Ile Ala Arg Ala
770 775 780

Leu Pro Gly Phe Pro Ile Leu Ala Thr Gly Gly Ile Asp Ser Ala Glu
785 790 795 800

Ser Gly Leu Gln Phe Leu His Ser Gly Ala Ser Val Leu Gln Val Cys
805 810 815

Ser Ala Ile Gln Asn Gln Asp Phe Thr Val Ile Glu Asp Tyr Cys Thr
820 825 830

Gly Leu Lys Ala Leu Leu Tyr Leu Lys Ser Ile Glu Glu Leu Gln Asp
835 840 845

Trp Asp Gly Gln Ser Pro Ala Thr Val Ser His Gln Lys Gly Lys Pro
850 855 860

Val Pro Arg Ile Ala Glu Leu Met Asp Lys Lys Leu Pro Ser Phe Gly
865 870 875 880

Pro Tyr Leu Glu Gln Arg Lys Lys Ile Ile Ala Glu Asn Lys Ile Arg
885 890 895

Leu Lys Glu Gln Asn Val Ala Phe Ser Pro Leu Lys Arg Asn Cys Phe
900 905 910

Ile Pro Lys Arg Pro Ile Pro Thr Ile Lys Asp Val Ile Gly Lys Ala
915 920 925

Leu Gln Tyr Leu Gly Thr Phe Gly Glu Leu Ser Asn Val Glu Gln Val
930 935 940

Val Ala Met Ile Asp Glu Glu Met Cys Ile Asn Cys Gly Lys Cys Tyr
945 950 955 960

Met Thr Cys Asn Asp Ser Gly Tyr Gln Ala Ile Gln Phe Asp Pro Glu
965 970 975

Thr His Leu Pro Thr Ile Thr Asp Thr Cys Thr Gly Cys Thr Leu Cys
980 985 990

Leu Ser Val Cys Pro Ile Val Asp Cys Ile Lys Met Val Ser Arg Thr
995 1000 1005

Thr Pro Tyr Glu Pro Lys Arg Gly Val Pro Leu Ser Val Asn Pro Val
1010 1015 1020

Cys
1025

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Asp Ala Leu Glu Leu Asn Leu Ser Cys
1 5

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Lys Asp Val Ala Asp Ile Glu
1 5

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 8

(D) OTHER INFORMATION: /product= "OTHER"

/note= "X = any amino acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Lys Ala Glu Ala Ser Gly Ala Xaa Ala Leu Glu Leu Asn Leu Ser Cys
1 5 10 15

Pro His Gly Met Gly Glu Arg
20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: modified_base

(B) LOCATION: one-of(6, 12, 15, 18)

(D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AARGGNGARG CNTCNGGNGC 20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: modified_base

(B) LOCATION: one-of(3, 9, 15)

(D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TCNCCCATNC CRTGNGG 17

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Lys Ala Glu Ala Ser Gly Ala
1 5

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Pro His Gly Met Gly Glu
1 5

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ACTCGATGCG ACATCGATTT TTTTTTTTTT TTTT 34

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGCCCTGGAG TTAAATTTAT CGTG 24

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Val Leu Ser Lys Asp Val Ala Asp Ile Glu Ser Ile Leu Ala Leu Asn
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: modified_base

(B) LOCATION: one-of(9, 12)

(D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = Inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AARGAYGTNG CNGATATCGA

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AACCCAGCGA CAGATGTTCC 20

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTCGTGTGCT TGATGTCATC 20

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCTTCTCGCA ATTAAAGCAG 20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCTCTGAAGG TTCCAGAATC GATAG 25

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CTGGAATTCG GCTTAAAGGA CGTGGCGG 28

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CTGGAATTCG GCTT 14

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Gly Leu Lys Ala Asp Gly Thr Pro Trp Pro Ala Val Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Ser Ile Leu Ala Leu Asn
1 5

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 1

(D) OTHER INFORMATION: /product= "OTHER"

/note= "X = any amino acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Xaa Ala Leu Glu Leu Asn Leu Ser Cys

1 5

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TGTAGGCACT GCCATGGCCC CTGTG

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TTCACAAATC ACCTTAACAC ACC 23

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTGGTGGTTT AAGTACTTCT GAAATTCC 28

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTTGCTCTGT CCGAACAAAC TGCATAGCA 29

CLAIMS

1. A DNA segment comprising an isolated mammalian dihydropyrimidine dehydrogenase (DPD) gene.

2. The DNA segment of claim 1, comprising an isolated bovine DPD gene.

3. The DNA segment of claim 2, comprising a bovine DPD gene that encodes a DPD protein or peptide that includes a contiguous amino acid sequence as set forth by a contiguous sequence from SEQ ID NO:2.

4. The DNA segment of claim 3, comprising a sequence region that consists of bovine DPD gene that includes a contiguous nucleic acid sequence as set forth by a contiguous sequence from the sequence between position 68 and position 3142 of SEQ ID NO:1.

5. The DNA segment of claim 3, comprising a bovine DPD gene that encodes a DPD peptide of from about 15 to about 150 amino acids in length.

6. The DNA segment of claim 3, comprising a bovine DPD gene that encodes a DPD protein of about 1025 amino acids in length.

7. The DNA segment of claim 6, comprising a bovine DPD gene that has a nucleic acid sequence as set forth by the sequence from position 68 to position 3142 of SEQ ID NO:1.

8. The DNA segment of claim 1, comprising an isolated human DPD gene.

9. The DNA segment of claim 8, comprising a human DPD gene that encodes a DPD protein or peptide that includes a contiguous amino acid sequence as set forth by a contiguous sequence from SEQ ID NO:4.

10. The DNA segment of claim 9, comprising a human DPD gene that includes a contiguous nucleic acid sequence as set forth by a contiguous sequence from the sequence between position 49 and position 3123 of SEQ ID NO:3.

11. The DNA segment of claim 9, comprising a human DPD gene that encodes a DPD peptide of from about 15 to about 150 amino acids in length.

12. The DNA segment of claim 9, comprising a human DPD gene that encodes a DPD protein of about 1025 amino acids in length.

13. The DNA segment of claim 12, comprising a human bovine DPD gene that has a nucleic acid sequence as set forth by the sequence from position 49 to position 3123 of SEQ ID NO:3.

14. The DNA segment of claim 1, positioned under the control of a promoter.

15. The DNA segment of claim 14, positioned under the control of a recombinant promoter.

16. The DNA segment of claim 1, further defined as a recombinant vector.

17. A nucleic acid segment that comprises at least a 14 nucleotide long contiguous stretch that corresponds to a nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3.

18. The nucleic acid segment of claim 17, further defined as comprising at least a 20 nucleotide long contiguous stretch that corresponds to a nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3.

19. The nucleic acid segment of claim 18, further defined as comprising at least a 30 nucleotide long contiguous stretch that corresponds to a nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3.

20. The nucleic acid segment of claim 19, further defined as comprising at least a 50 nucleotide long contiguous stretch that corresponds to a nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3.

21. The nucleic acid segment of claim 20, further defined as comprising at least a 100 nucleotide long contiguous stretch that corresponds to a nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3.

22. The nucleic acid segment of claim 21, further defined as comprising at least a 200 nucleotide long contiguous stretch that corresponds to the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3.

23. The nucleic acid segment of claim 22, further defined as comprising a 4414 nucleotide long contiguous stretch that corresponds to the nucleic acid sequence of SEQ ID NO:1.

24. The nucleic acid segment of claim 22, further defined as comprising a 4368 nucleotide long contiguous stretch that corresponds to the nucleic acid sequence of SEQ ID NO:3.

25. The nucleic acid segment of claim 17, further defined as comprising a nucleic acid fragment of up to 10,000 basepairs in length.

26. The nucleic acid segment of claim 25, further defined as comprising a nucleic acid fragment of up to 5,000 basepairs in length.

27. The nucleic acid segment of claim 26, further defined as comprising a nucleic acid fragment of up to 3,000 basepairs in length.

28. The nucleic acid segment of claim 27, further defined as comprising a nucleic acid fragment of up to 1,000 basepairs in length.

29. The nucleic acid segment of claim 17, further defined as a DNA segment.

30. A recombinant host cell comprising a DNA segment that comprises an isolated mammalian DPD gene.

31. The recombinant host cell of claim 30, further defined as a prokaryotic host cell.

32. The recombinant host cell of claim 31, further defined as *E. coli*.

33. The recombinant host cell of claim 30, further defined as a eukaryotic host cell.

34. The recombinant host cell of claim 33, further defined as an insect cell or a yeast cell.

35. The recombinant host cell of claim 30, wherein the DNA segment is introduced into the cell by means of a recombinant vector.

36. The recombinant host cell of claim 35, wherein the host cell expresses the DNA segment to produce the encoded protein or peptide.

37. The recombinant host cell of claim 36, wherein the expressed protein or peptide includes an amino acid sequence essentially as set forth by a contiguous sequence from SEQ ID NO:2 or SEQ ID NO:4.

38. The recombinant host cell of claim 37, wherein the expressed protein has an amino acid sequence essentially as set forth by a contiguous sequence from SEQ ID NO:2.

39. The recombinant host cell of claim 37, wherein the expressed protein has an amino acid sequence essentially as set forth by a contiguous sequence from SEQ ID NO:4.

40. A method of using a DNA segment that includes an isolated mammalian DPD gene, comprising:

(a) preparing a recombinant vector in which a DPD-encoding DNA segment is positioned under the control of a promoter;

(b) introducing said recombinant vector into a recombinant host cell;

(c) culturing the recombinant host cell under conditions effective to allow expression of an encoded DPD protein or peptide; and

(d) collecting said expressed DPD protein or peptide.

41. A method for detecting DPD in a sample, comprising obtaining sample nucleic acids from a sample suspected of containing DPD, contacting said sample nucleic acids with a nucleic acid segment that encodes a mammalian DPD protein or peptide under conditions effective to allow hybridization of substantially complementary nucleic acids, and detecting the hybridized complementary nucleic acids thus formed.

42. The method of claim 41, wherein the sample nucleic acids contacted are located within a cell.

43. The method of claim 41, wherein the sample nucleic acids are separated from a cell prior to contact.

44. The method of claim 41, wherein the sample nucleic acids are DNA.

45. The method of claim 41, wherein the sample nucleic acids are RNA.

46. The method of claim 41, wherein the nucleic acid segment comprises a detectable label and the hybridized complementary nucleic acids are detected by detecting said label.

47. The method of claim 46, wherein the nucleic acid segment comprises a radio, enzymatic or fluorescent label.

48. The method of claim 41, wherein the sample suspected of containing DPD is a biological sample obtained from a patient suspected of having cancer.

49. A DPD detection kit comprising, in suitable container means, a nucleic acid segment that encodes a mammalian DPD protein or peptide and a detection reagent.

50. The DPD detection kit of claim 49, wherein said detection reagent is a detectable label that is linked to said nucleic acid segment.

51. The DPD detection kit of claim 49, further comprising, in a suitable container, a restriction enzyme.

52. A method for determining a therapeutically effective dose of 5-fluorouracil (FUra) for administration to a patient, comprising determining the amount of DPD present within a biological sample from said patient and adjusting the dose of FUra to be administered according to the amount of DPD detected.

53. The method of claim 52, wherein the amount of DPD present within said biological sample is determined by means of an immunoassay to detect a DPD protein.

54. The method of claim 52, wherein the amount of DPD present within said biological sample is determined by means of a molecular biological assay to detect a DPD nucleic acid segment.

55. The method of claim 52, wherein said biological sample is a blood sample.

56. The method of claim 52, wherein upon detecting an increased amount of DPD within said sample the dose of FUra to be administered is increased.

57. The method of claim 52, wherein upon detecting an increased amount of DPD within said sample the FUra is combined with an inhibitor of DPD prior to administration.

58. The method of claim 57, wherein upon detecting an increased amount of DPD within said sample the FUra is combined with uridine, 5-ethynyluracil (EU), interferon, leucovorin, cimetidine (CMT) or 5-benzyloxybenzyluracil (BBU) prior to administration.

59. The method of claim 52, wherein upon detecting a decreased amount of DPD within said sample the dose of FUra to be administered is decreased.

60. The method of claim 52, wherein upon detecting a significantly decreased amount of DPD within said sample the FUra is combined with a pharmaceutically-acceptable composition comprising DPD peptide prior to administration.

61. The method of claim 52, wherein upon detecting a significantly decreased amount of DPD within said sample the FUra is administered after administration of a pharmaceutically-acceptible composition comprising a DPD-encoding DNA segment.

62. The method of claim 52, wherein upon detecting a significantly decreased amount of DPD within said sample, no FUra is administered.

63. A monoclonal antibody that has binding affinity for human DPD.

64. The monoclonal antibody of claim 63, obtainable by the method of immunizing an animal with recombinant human DPD in an amount effective to stimulate the generation of B cells producing antibodies specific for DPD, immortalizing said B cells and obtaining a monoclonal antibody therefrom.

65. A method for diagnosing DPD deficiency in a human, comprising determining the presence of a frameshift mutation in a DPD-encoding DNA segment present within a biological sample from a patient suspected of having DPD deficiency, wherein the presence of a frameshift mutation in a DPD-encoding DNA segment results in a decreased amount of active DPD polypeptide, in comparison to the amount within a sample from a normal subject, and said frameshift mutation is indicative of a patient with DPD deficiency.

66. The method of claim 65, wherein the frameshift mutation in said DPD-encoding DNA segment is determined by means of a molecular biological assay to detect the deletion of an adenosine residue at codon 318 within said DPD-encoding nucleic acid.

67. The method of claim 65, wherein said frameshift mutation in said DPD-encoding DNA segment results in multiple incorrect codons and a premature stop at codon 335 within said DPD-encoding nucleic acid.

68. The method of claim 65, wherein said frameshift mutation in said DPD-encoding DNA segment results in a truncated DPD having a molecular weight of approximately 40 kDa.

69. The method of claim 65, wherein said biological sample is a blood sample.

70. The method of claim 65, further defined as a method for diagnosing increased sensitivity to FUra.

71. The method of claim 65, wherein said frameshift mutation is identified by sequencing said DNA segment.

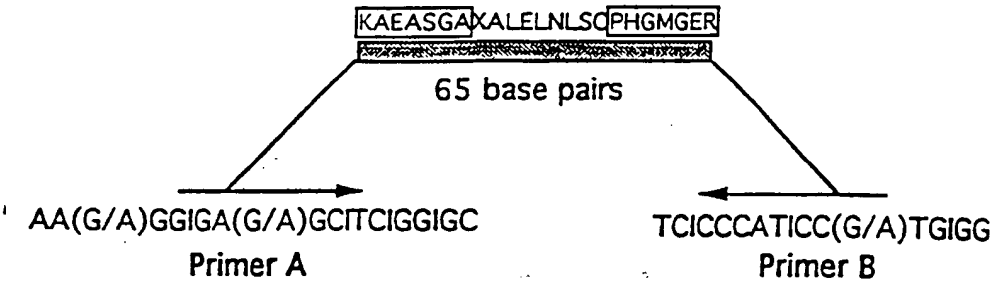
KAEASGA
XALELNLSC
PHGMGER
65 base pairs
AA(G/A)GGIGA(G/A)GCITCIGGIGC
Primer A
TCICCCATICC(G/A)TGIGG
Primer B

FIG. 1A

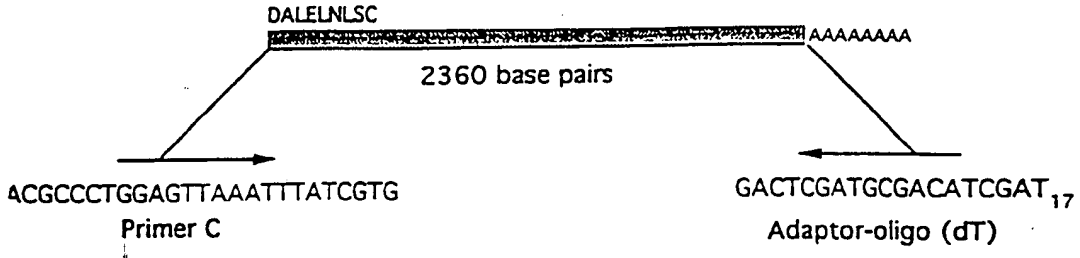
DALELNLSC
AAAAAAAA
2360 base pairs
ACGCCCTGGAGTTAAATTTATCGTG
Primer C
GACTCGATGCGACATCGAT17
Adaptor-oligo (dT)

FIG. 1B

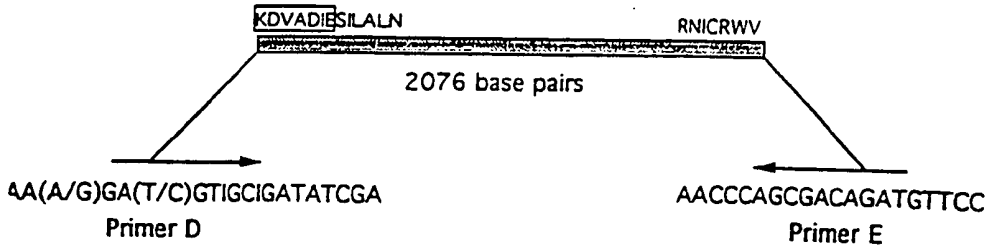
KDVADIESILALN
RNICRWV
2076 base pairs
4A(A/G)GA(T/C)GTIGCIGATATCGA
Primer D
AACCCAGCGACAGATGTTCC
Primer E

FIG. 1C

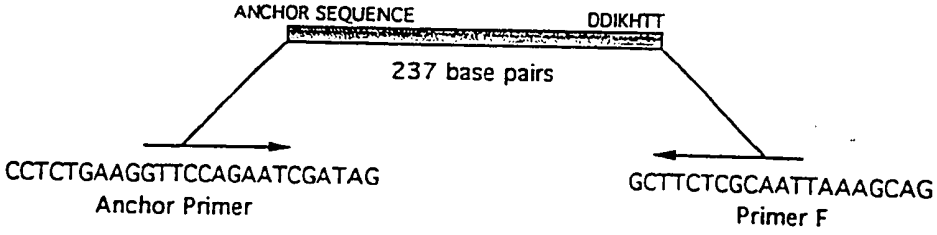
ANCHOR SEQUENCE
DDIKHTT
237 base pairs
CCTCTGAAGGTTCCAGAATCGATAG
Anchor Primer
GCTTCTCGCAATTAAAGCAG
Primer F

FIG. 1D

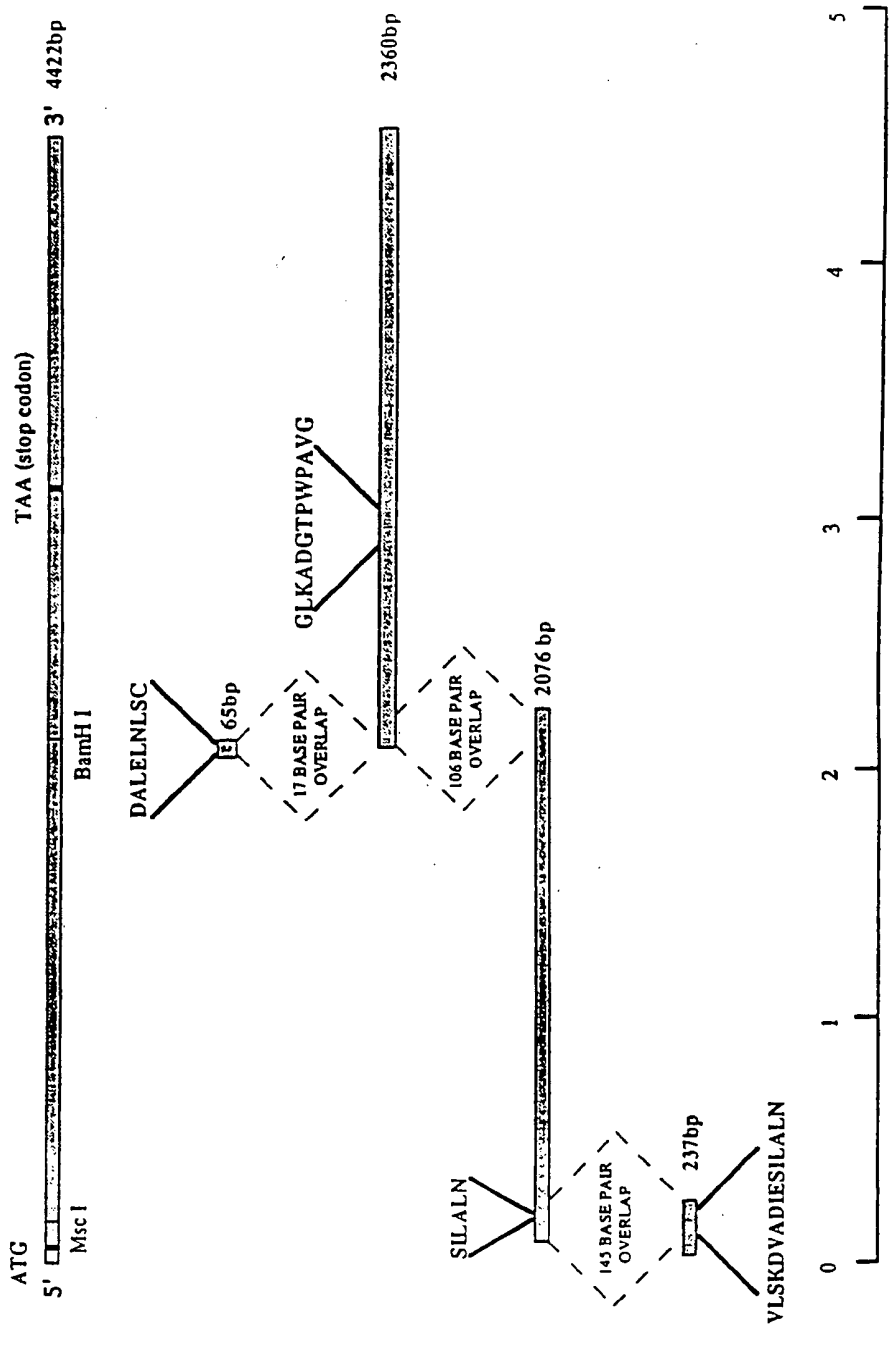

FIG. 2

FIG. 3

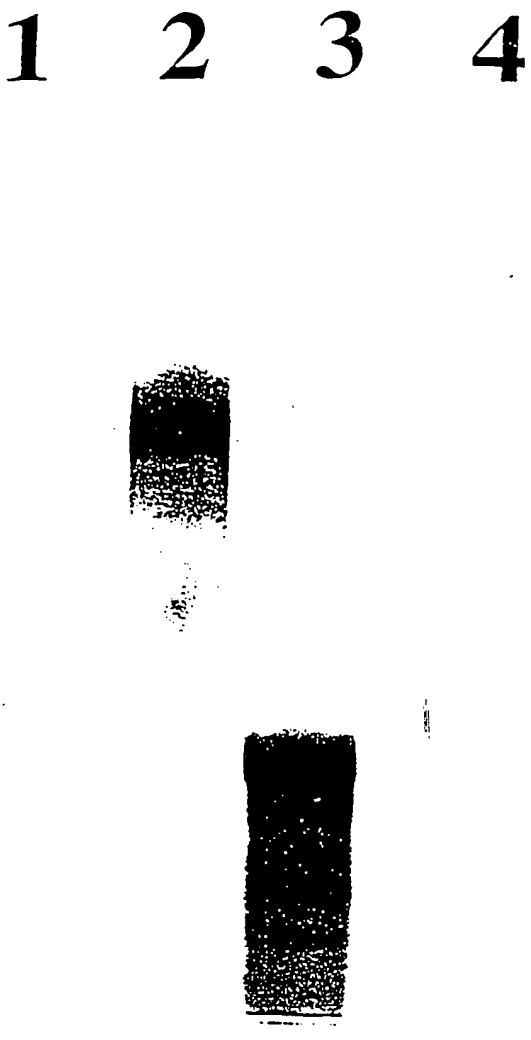

FIG. 4

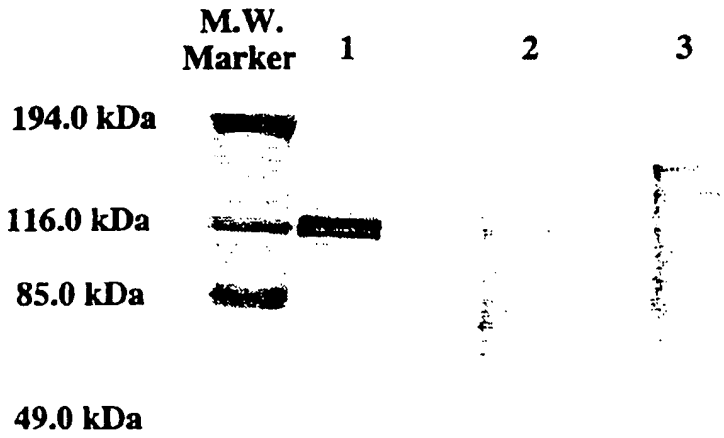

FIG. 5

```
1          ACTTTCGCTGAAGCCTGAGGACGCGGAAGGGTTCGTGGCAAGGAAACCCCAGGCTCTGGGCATCGCC  67

68    ATGGCCCCTGTGCTAAGCAAGGACGTGGCGGATATCGAGAGTATCCTGGCTTTAAATCCTCGAACACAGTCTCGTGCAACTCTGCGTTCC  157
1      M  A  P  V  L  S  K  D  V  A  D  I  E  S  I  L  A  L  N  P  R  T  Q  S  R  A  T  L  R  S  30

158   ACTTTGGCCAAGAAATTAGACAAGAAACATTGGAAAAGAAATCCTGATAAGAACTGCTTTAATTGCGAGAAGCTGGAGAATAATTTTGAT  247
31     T  L  A  K  K  L  D  K  K  H  W  K  R  N  P  D  K  N  C  F  N  C  E  K  L  E  N  N  F  D  60

248   GACATCAAGCACACGACTCTTGGTGAGCGAGGAGCTCTCCGAGAAGCAATGAGATGCCTGAAATGTGCAGATGCCCCCTGTCAGAAGAGC  337
61     D  I  K  H  T  T  L  G  E  R  G  A  L  R  E  A  M  R  C  L  K  C  A  D  A  P  C  Q  K  S  90

338   TGTCCAACAAATCTAGATATCAAATCGTTCATCACAAGTATCTCAAACAAGAACTATTATGGAGCTGCTAAGATGATATTTTCTGACAAC  427
91     C  P  T  N  L  D  I  K  S  F  I  T  S  I  S  N  K  N  Y  Y  G  A  A  K  M  I  F  S  D  N  120

428   CCACTTGGTCTGACCTGTGGAATGGTATGCCCGACCTCTGATCTTTGTGTAGGTGGATGTAATTTGTATGCCACTGAAGAGGGACCAATT  517
121    P  L  G  L  T  C  G  M  V  C  P  T  S  D  L  C  V  G  G  C  N  L  Y  A  T  E  E  G  P  I  150

518   AATATTGGTGGATTGCAGCAATATGCTACTGAGGTATTCAAAGCAATGAATATCCCACAAATCAGGAATCCTTCTCTGCCTCCCCCAGAA  607
151    N  I  G  G  L  Q  Q  Y  A  T  E  V  F  K  A  M  N  I  P  Q  I  R  N  P  S  L  P  P  P  E  180

608   AAAATGCCTGAAGCCTATTCTGCAAAGATTGCTCTTTTGGGTGCTGGGCCTGCAAGTATAAGTTGTGCTTCGTTCTTGGCTCGATTAGGC  697
181    K  M  P  E  A  Y  S  A  K  I  A  L  L  G  A  G  P  A  S  I  S  C  A  S  F  L  A  R  L  G  210

698   TACAATGACATCACTATATTTGAAAAACAAGAATACGTTGGTGGTATAAGTACTTCTGAAATCCCTCAGTTCCGGCTGCCATATGATGTA  787
211    Y  N  D  I  T  I  F  E  K  Q  E  Y  V  G  G  I  S  T  S  E  I  P  Q  F  R  L  P  Y  D  V  240

788   GTAAATTTTGAGATTGAGCTTATGAAGGACCTTGGTGTAAAGATAATTTGTGGTAAAAGCCTTTCAGTGAATGACATTACTCTTAGTACT  877
241    V  N  F  E  I  E  L  M  K  D  L  G  V  K  I  I  C  G  K  S  L  S  V  N  D  I  T  L  S  T  270

878   TTGAAAGAAGAAGGGTACAAAGCTGCTTTCATTGGGATAGGTTTGCCAGAACCCAAGAAGGATCACATCTTCCAAGGCCTGACACAGGAC  967
271    L  K  E  E  G  Y  K  A  A  F  I  G  I  G  L  P  E  P  K  K  D  H  I  F  Q  G  L  T  Q  D  300

968   CAGGGGTTTTACACTTCCAAAGACTTTCTGCCTCTTGTAGCCAAAAGCAGTAAAGCAGGAATGTGCGCCTGTCACTCTCCATTGCTGTCG  1057
301    Q  G  F  Y  T  S  K  D  F  L  P  L  V  A  K  S  S  K  A  G  M  C  A  C  H  S  P  L  L  S  330

1058  ATACGGGGAACCGTGATTGTACTCGGAGCTGGAGACACAGCTTTCGACTGTGCAACATCCGCTTTACGTTGTGGAGCCCGCCGAGTGTTC  1147
331    I  R  G  T  V  I  V  L  G  A  G  D  T  A  F  D  C  A  T  S  A  L  R  C  G  A  R  R  V  F  360

1148  ATCGTCTTCAGAAAAGGCTTTGTTAATATAAGAGCTGTCCCTGAGGAGGTGGAGCTTGCTAGAGAAGAAAAATGTGAATTTTTGCCTTTC  1237
361    I  V  F  R  K  G  F  V  N  I  R  A  V  P  E  E  V  E  L  A  R  E  E  K  C  E  F  L  P  F  390

1238  TTGTCTCCACGGAAGGTTATAGTAAAAGGTGGGAGAATTGTTGCCATGCAATTTGTTCGGACAGAGCAAGATGAAACTGGAAAATGGAAT  1327
391    L  S  P  R  K  V  I  V  K  G  G  R  I  V  A  M  Q  F  V  R  T  E  Q  D  E  T  G  K  W  N  420

1328  GAAGATGGAGATCAGATAGCCTGTCTGAAAGCCGATGTGGTCATCAGTGCCTTTGGCTCAGTTCTGAGTGATCCTAAAGTAAAAGAAGCC  1417
421    E  D  G  D  Q  I  A  C  L  K  A  D  V  V  I  S  A  F  G  S  V  L  S  D  P  K  V  K  E  A  450
```

FIG. 6A

```
1418  TTGAGCCCTATAAAATTTAACAGATGGGATCTCCCAGAAGTAGATCCAGAAACCAGTGCAAACCATGCAAACCAGTGAGCCATGGGTGTTTGCAGGTGGT  1507
 451  L  S  P  I  K  F  N  R  W  D  L  P  E  V  D  P  E  T  M  Q  T  S  E  P  W  V  F  A  G  G              480

1508  GATGTGGTTGGTATAGCCAACACTACAGTGGAAGCCGTGAATGATGGAAAGCAAGCCTCTTGGTACATTCACAGATATATACAGTCACAA            1597
 481  D  V  V  G  I  A  N  T  T  V  E  A  V  N  D  G  K  Q  A  S  W  Y  I  H  R  Y  I  Q  S  Q              510

1598  TATGGAGCTTCAGTTTCTGCTAAGCCCGAACTCCCCCCTGTTTTATACTCCCATTGATCTGGTGGACATTAGTGTGGAAATGGCTGCATTG            1687
 511  Y  G  A  S  V  S  A  K  P  E  L  P  L  F  Y  T  P  I  D  L  V  D  I  S  V  E  M  A  A  L              540

1688  AAGTTTACAAATCCTTTGGTCTTGCCAGTGCAACTCCCAACTACCAGTTCGTCAATGATTCGAAGAGCTTTTGAAGCTGGATGGGCCTTT            1777
 541  K  F  T  N  P  F  G  L  A  S  A  T  P  T  T  S  S  M  I  R  R  A  F  E  A  G  W  A  F              570

1778  GCTCTGACCAAAACTTTCTCTCTTGATAAGGATATAGTTGACAAATGTTTCACCCAGAATCATCCGGGGACCACCTCTGGCCCCATGTAT            1867
 571  A  L  T  K  T  F  S  L  D  K  D  I  V  T  N  V  S  P  R  I  I  R  G  T  T  S  G  P  M  Y              600

1868  GGCCCTGGACAAAGCTCTTTCCTGAATATTGAGCTCATCAGTGAAAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAAGCC            1957
 601  G  P  G  Q  S  S  F  L  N  I  E  L  I  S  E  K  T  A  A  Y  W  C  Q  S  V  T  E  L  K  A              630

1958  GACTTTCCAGACAATATTGTGATTGCTAGCATCATGTGCAGTTACAACAGAAATGACTGGATGGAACTCTCCAGAAAGGCTGAGGCTTCT            2047
 631  D  F  P  D  N  I  V  I  A  S  I  M  C  S  Y  N  R  N  D  W  M  E  L  S  R  K  A  E  A  S              660

2048  GGAGCAGACGCCCTGGAGTTAAATTTATCGTGTCCGCATGGCATGGGAGAAGAGGAATGGGTCTGGCTTGTGGACAGGATCCAGAGCTG            2137
 661  G  A  D  A  L  E  L  N  L  S  C  P  H  G  M  G  E  R  G  M  G  L  A  C  G  Q  D  P  E  L              690

2138  GTGCGGAACATCTGTCGCTGGGTTAGGCAAGCTGTTCGGATTCCTTTTTTTGCCAAGTTGACCCCAAATGTCACTGATATTGTAAGCATA            2227
 691  V  R  N  I  C  R  W  V  R  Q  A  V  R  I  P  F  F  A  K  L  T  P  N  V  T  D  I  V  S  I              720

2228  GCCAGAGCTGCAAAGGAAGGTGGGGCAAATGGTGTTACAGCTACCAACACTGTCTCAGGTCTCATGGGATTAAAAGCTGATGGCACACCC            2317
 721  A  R  A  A  K  E  G  G  A  N  G  V  T  A  T  N  T  V  S  G  L  M  G  L  K  A  D  G  T  P              750

2318  TGGCCAGCAGTGGGCCGTGAGAAGCGGACTACATACGGCGGAGTGTCCGGCACAGCCATCAGACCTATTGCTTTGAGAGCTGTGACCACC            2407
 751  W  P  A  V  G  R  E  K  R  T  T  Y  G  G  V  S  G  T  A  I  R  P  I  A  L  R  A  V  T  T              780

2408  ATTGCTCGTGCTTTGCCTGAATTTCCCATTTGGCCACTGGTGGAATTGACTCAGCTGAAAGTGGTCTTCAGTTTCTCCACGGTGGTGCT            2497
 781  I  A  R  A  L  P  E  F  P  I  L  A  T  G  G  I  D  S  A  E  S  G  L  Q  F  L  H  G  G  A              810

2498  TCGGTGCTCCAGGTATGCAGTGCTATTCAAAATCAGGATTTCACTATCATCCAAGACTACTGCACTGGCCTCAAAGCCTTGCTTTATCTG            2587
 811  S  V  L  Q  V  C  S  A  I  Q  N  Q  D  F  T  I  I  Q  D  Y  C  T  G  L  K  A  L  L  Y  L              840

2588  AAAAGCATTGAAGAACTACAGGACTGGGATGGGCAGAGTCCAGCCACGAAGAGTCACCAGAAAGGGAAACCAGTCCCTTGTATTGCTGAA            2677
 841  K  S  I  E  E  L  Q  D  W  D  G  Q  S  P  A  T  K  S  H  Q  K  G  K  P  V  P  C  I  A  E              870

2678  CTTGTGGGAAAGAAACTGCCAAGCTTTGGACCTTATCTTGAGAAGTGCAAGAAATCATAGCAGAGGAAAAGTTGAGACTGAAAAAAGAA            2767
 871  L  V  G  K  K  L  P  S  F  G  P  Y  L  E  K  C  K  K  I  I  A  E  E  K  L  R  L  K  K  E              900
```

FIG. 6B

```
2768  AATGTGACCGTTCTACCACTTGAAAGAAACCATTTTATCCCAAAAAAACCTATTCCTTCTGTTAAGGATGTGATTGGAAAAGCTCTGCAG   2857
 901  N  V  T  V  L  P  L  E  R  N  N  H  F  I  P  K  K  P  I  P  S  V  K  D  V  I  G  K  A  L  Q   930

2858  TACCTTGGAACATATGGTGAACTGAACAACACAGAGCAGGTTGTGGCTGTGATCGATGAAGAGATGTGTATCAACTGTGGCAAATGCTAC   2947
 931  Y  L  G  T  Y  G  E  L  N  N  T  E  Q  V  V  A  V  I  D  E  E  M  C  I  N  C  G  K  C  Y   960

2948  ATGACCTGTAATGACTCTGGCTACCAGGCTATCCAGTTTGATCCTGAAACCCACCTGCCCACCGTTACTGACACTTGTACAGGCTGTACC   3037
 961  M  T  C  N  D  S  G  Y  Q  A  I  Q  F  D  P  E  T  H  L  P  T  V  T  D  T  C  T  G  C  T   990

3038  CTGTGTCTCTCCGTCTGCCCTATTATCGACTGCATCAAAATGGTTTCCAGGACAACACCTTATGAACCAAGAGAGGCTTGCCCTTGGCT   3127
 991  L  C  L  S  V  C  P  I  I  D  C  I  K  M  V  S  R  T  T  P  Y  E  P  K  R  G  L  P  L  A   1020

3128  GTGAATCCTGTGTCTTAAGGTGATTTGTGAAACAGTTGCAGTGAACTTCGAGGTCACCTACTTATGCTGATCTTTTCAATAGTGATCATT   3217
1021  V  N  P  V  S    1025

3218  ATGCTCAGCTTTTTCTAAATTCAAACATATAATTTCTAAATTTAAAAAAAGATTATTTCTAAAGAAATTTCTAAATTTTAAAAATGTCTG   3307

3308  CTTTCAGTGATCATTCAATTAATGGTCATAAAATAGAATAATTCTTTTCTGAGCAGAATTGTTCAATATAACTATGGAGCAGTTAATTGG   3397

3398  ATGTTCACCATCAGTTGTCCATTATGAAAAAATTAACTTTTTTGTAGCAATTAATGCTACACTTTCAAATTGCCCTATGCCGAGTTCTG   3487

3488  TCTTTGATTTCTAATTGTAAGGGAAATTAAGTATTTTAGAACAAAGTACAATTTAACTTACAACAAATGTTTCCAAGGAAACATTTTATA   3577

3578  ATTAAAAATTACATTTAATTTAACTCTGTTTCTAAGCAAAAGTAATTAGCTCCATAAAGCTCAGATGAAGTCAAATAATTATTTACTGT   3667

3668  GGTAGCAAAAGAAAGCCAATGAGGGTTTGGGAAACTTTCCTTAAGGTCTCTTCACTGAAATAACTGGATACTGAAGGCGAGAGTGTTCAG   3757

3758  TAACCATTTGTATCAAGCTATGCTATTCACCGCTCAGGCCTGAGATGTGTGTCCAAATGCTACCAATGAATCAACATGACATTCCTGTTT   3847

3848  AAATATTTAAACTATGTTCCTAACAAAGTAAGACATTAGGATGGAACTCTGGTTAAAGCCACTCTTTTGCTGTGCACAGATCTGTTCTAT   3937

3938  CTGCTTCTAATATAGTCACCTTCGTGATCCTAGCAATTAATGTTTGAACACAGCACAGATTATACAGAAGTGGGGTCATGTGCTTCTTTA   4027

4028  TTCAAGAATGAGAAATCCAGTATGGGTAATATATATTATATAGGGTGATACCACTTTACCAACTCTTTATTTTAGTGTCCATGTTGAATTT   4117

4118  CGAAAGTGATTAAAAAAGAAATGGTATTTTCTGTTACTGCCAAATAATATTTTATATTCCTCGATTTTTAAAATCAGCAAATAGCATCT   4207

4208  TATAAACTTGTTTATCTCTTCTTTGTGGCATATTTTAATATGAATCCATAAGTAGTAAATCTTCATGTAATCATCCATAGCACCTTTCTA   4297

4298  TGACAAATGCAAGATCAAGAGAAAAATAAATGTTTGATTATGCACTTTTAGAAATGCACATTTACCACACAAAATCTGTATGATCAAATAAT   4387

4388  ATTAAATAAAATTTTATAAAGCATTTTAAAAAAAAAAAAAAAAA   4422
```

FIG. 6C

```
                                      ACGCAAGGAGGGTTTGTCACTGGCAGACTCGAGACTGTAGGCACTGCC   48
  49 ATGGCCCCTGTGCTCAGTAAGGACTCGGCGGACATCGAGAGTATCCTGGCTTTAAATCCTCGAACACAAACTCATGCAACTCTGTGTTCC  138
   1  M  A  P  V  L  S  K  D  S  A  D  I  E  S  I  L  A  L  N  P  R  T  Q  T  H  A  T  L  C  S   30

 139 ACTTCGGCCAAGAAATTAGACAAGAAACATTGGAAAAGAAATCCTGATAAGAACTGCTTTAATTGTGAGAAGCTGGAGAATAATTTTGAT  228
  31  T  S  A  K  K  L  D  K  K  H  W  K  R  N  P  D  K  N  C  F  N  C  E  K  L  E  N  N  F  D   60

 229 GACATCAAGCACACGACTCTTGGTGAGCGAGGAGCTCTCCGAGAAGCAATGAGATGCCTGAAATGTGCAGATGCCCCGTGTCAGAAGAGC  318
  61  D  I  K  H  T  T  L  G  E  R  G  A  L  R  E  A  M  R  C  L  K  C  A  D  A  P  C  Q  K  S   90

 319 TGTCCAACTAATCTTGATATTAAATCATTCATCACAAGTATTGCAAACAAGAACTATTATGGAGCTGCTAAGATGATATTTTCTGACAAC  408
  91  C  P  T  N  L  D  I  K  S  F  I  T  S  I  A  N  K  N  Y  Y  G  A  A  K  M  I  F  S  D  N  120

 409 CCACTTGGTCTGACTTGTGGAATGGTATGTCCAACCTCTGATCTTTGTGTAGGTGGATGCAATTTATATGCCACTGAAGAGGGACCCATT  498
 121  P  L  G  L  T  C  G  M  V  C  P  T  S  D  L  C  V  G  G  C  N  L  Y  A  T  E  E  G  P  I  150

 499 AATATTGGTGGATTGCAGCAATTTGCTACTGAGGTATTCAAAGCAATGAGTATCCCACAGATCAGAAATCCTTCGCTGCCTCCCCCAGAA  588
 151  N  I  G  G  L  Q  Q  F  A  T  E  V  F  K  A  M  S  I  P  Q  I  R  N  P  S  L  P  P  P  E  180

 589 AAAATGTCTGAAGCCTATTCTGCAAAGATTGCTCTTTTTGGTGCTGGGCCTGCAAGTATAAGTTGTGCTTCCTTTTTGGCTCGATTGGGG  678
 181  K  M  S  E  A  Y  S  A  K  I  A  L  F  G  A  G  P  A  S  I  S  C  A  S  F  L  A  R  L  G  210

 679 TACTCTGACATCACTATATTTGAAAAACAAGAATATGTTGGTGGTTTAAGTACTTCTGAAATTCCTCAGTTCCGGCTGCCGTATGATGTA  768
 211  Y  S  D  I  T  I  F  E  K  Q  E  Y  V  G  G  L  S  T  S  E  I  P  Q  F  R  L  P  Y  D  V  240

 769 GTGAATTTTGAGATTGAGCTAATGAAGGACCTTGGTGTAAAGATAATTTGCGGTAAAAGCCTTTCAGTGAATGAAATGACTCTTAGCACT  858
 241  V  N  F  E  I  E  L  M  K  D  L  G  V  K  I  I  C  G  K  S  L  S  V  N  E  M  T  L  S  T  270

 859 TTGAAAGAAAAAGGCTACAAAGCTGCTTTCATTGGAATAGGTTTGCCAGAACCCAATAAAGATGCCATCTTCCAAGGCCTGACGCAGGAC  948
 271  L  K  E  K  G  Y  K  A  A  F  I  G  I  G  L  P  E  P  N  K  D  A  I  F  Q  G  L  T  Q  D  300
                                                       ↓
 949 CAGGGGTTTTATACATCCAAAGACTTTTTGCCACTTGTAGCCAAAGGCAGTAAAGCAGGAATGTGCGCCTGTCACTCTCCATTGCCATCG 1038
 301  Q  G  F  Y  T  S  K  D  F  L  P  L  V  A  K  G  S  K  A  G  M  C  A  C  H  S  P  L  P  S  330

1039 ATACGGGGAGTCGTGATTGTACTTGGAGCTGGAGACACTGCCTTTGACTGTGCAACATCTGCTCTACGTTGTGGAGCTCGCCGTGTGTTC 1128
 331  I  R  G  V  V  I  V  L  G  A  G  D  T  A  F  D  C  A  T  S  A  L  R  C  G  A  R  R  V  F  360
```

FIG. 7A

```
1129 ATCGTCTTCAGAAAGGCTTTGTTAATATAAGAGCTGTCCCTGAGGAGATGGAACTTGCTAAGGAAGAAAGTGTGAATTTCTGCCATTC 1218
 361  I  V  F  R  K  G  F  V  N  I  R  A  V  P  E  E  M  E  L  A  K  E  E  K  C  E  F  L  P  F  390

1219 CTGTCCCCACGGAAGGTTATAGTAAAAGGTGGGAGAATTGTTGCTATGCAGTTTGTTCGGACAGAGCAAGTGAAACTGGAAAATGGAAT 1308
 391  L  S  P  R  K  V  I  V  K  G  G  R  I  V  A  H  Q  F  V  R  T  E  Q  D  E  T  G  K  W  N  420

1309 GAAGATGAAGATCAGATGGTCCATCTGAAAGCCGATGTGGTCATCAGTGCCTTTGGTTCAGTTCTGAGTGATCCTAAAGTAAAAGAAGCC 1398
 421  E  D  E  D  Q  M  V  H  L  K  A  D  V  V  I  S  A  F  G  S  V  L  S  D  P  K  V  K  E  A  450

1399 TTGAGCCCTATAAAATTTAACAGATGGGGTCTCCCAGAAGTAGATCCAGAAACTATGCAAACTAGTGAAGCATGGGTATTTGCAGGTGGT 1488
 451  L  S  P  I  K  F  N  R  W  G  L  P  E  V  D  P  E  T  H  Q  T  S  E  A  W  V  F  A  G  G  480

1489 GATGTCGTTGGTTTGGCTAACACTACAGTGGAATCGGTGAATGATGGAAAGCAAGCTTCTTGGTACATTCACAAATACGTACAGTCACAA 1578
 481  D  V  V  G  L  A  N  T  T  V  E  S  V  N  D  G  K  Q  A  S  W  Y  I  H  K  Y  V  Q  S  Q  510

1579 TATGGAGCTTCCGTTCTGCCAAGCCTGAACTACCCCTCTTTACACTCCTATTGATCTGGTGGACATTAGTGTAGAAATGGCCGGATTG 1668
 511  Y  G  A  S  V  S  A  K  P  E  L  P  L  F  Y  T  P  I  D  L  V  D  I  S  V  E  M  A  G  L  540

1669 AAGTTTATAAATCCTTTTGGTCTTGCTAGCGCAACTCCAGCCACCAGCACATCAATGATTCGAAGAGCTTTTGAAGCTGGATGGGGTTTT 1758
 540  K  F  I  N  P  F  G  L  A  S  A  T  P  A  T  S  T  S  H  I  R  R  A  F  E  A  G  W  G  F  570

1759 GCCCTCACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCGGGGAACCACCTCTGGCCCCATGTAT 1848
 571  A  L  T  K  T  F  S  L  D  K  D  I  V  T  N  V  S  P  R  I  I  R  G  T  T  S  G  P  M  Y  600

1849 GGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCT 1938
 601  G  P  G  Q  S  S  F  L  N  I  E  L  I  S  E  K  T  A  A  Y  W  C  Q  S  V  T  E  L  K  A  630

1939 GACTTTCCAGACAACATTGTGATTGCTAGCATTATGTGCAGTTACAATAAAAATGACTGGACGGAACTTGCCAAGAAGTCTGAGGATTCT 2028
 631  D  F  P  D  N  I  V  I  A  S  I  M  C  S  Y  N  K  N  D  W  T  E  L  A  K  K  S  E  D  S  660

2029 GGAGCAGATGCCCTGGAGTTAAATTTATCATGTCCACATGGCATGGGAGAAGAGGAATGGGCCTGGCCTGTGGGCAGGATCCAGAGCTG 2118
 661  G  A  D  A  L  E  L  N  L  S  C  P  H  G  M  G  E  R  G  H  G  L  A  C  G  Q  D  P  E  L  690

2119 GTGCGGAACATCTGCCGCTGGGTTAGGCAAGCTGTTCAGATTCCTTTTTTGCCAAGCTGACCCCAAATGTCACTGATATTGTGAGCATC 2208
 691  V  R  H  I  C  R  W  V  R  Q  A  V  Q  I  P  F  F  A  K  L  T  P  N  V  T  D  I  V  S  I  720

2209 GCAAGAGCTGCAAAGGAAGGTGGTGCCAATGGCGTTACAGCCACCAACACTGTCTCAGGTCTGATGGGATTAAAATCTGATGGCACACCT 2298
 721  A  R  A  A  K  E  G  G  A  N  G  V  T  A  T  N  T  V  S  G  L  M  G  L  K  S  D  G  T  P  750
```

FIG. 7B

```
2299 TGGCCAGCAGTGGGGATTGCAAAGCGAACTACATATGGAGGAGTGTCTGGGACAGCAATCAGACCTATTGCTTTGAGAGCTGTGACCTCC 2388
  751 W P A V G I A K K R T T Y G G G V S G T A I R P I A L R A V T S  780

2389 ATTGCTCGTGCTCTGCCTGGATTTCCCATTTTGGCTACTGGTGGAATTGACTCTGCTGAAAGTGGTCTTCAGTTTCTCCATAGTGGTGCT 2478
  781 I A R A L P G F P I L A T G G I D S A E S G L Q F L H S G A  810

2479 TCCGTCCTCCAGGTATGCAGTGCCATTCAGAATCAGGATTTCACTGTGATCGAAGACTACTGCACTGGCCTCAAAGCCCTGCTTTATCTG 2568
  811 S V L Q V C S A I Q N Q D F T V I E D Y C T G L K A L L Y L  840

2569 AAAAGCATTGAAGAACTACAAGACTGGGATGGACAGAGTCCAGCTACTGTGAGTCACCAGAAAGGGAAACCAGTTCCACGTATAGCTGAA 2658
  841 K S I E E L Q D W D G Q S P A T V S H Q K G K P V P R I A E  870

2659 CTCATGGACAAGAAACTGCCAAGTTTTGGACCTTATCTGGAACAGCGCAAGAAAATCATAGCAGAAAACAAGATTAGACTGAAAGAACAA 2748
  871 L M D K K L P S F G P Y L E Q R K K I I A E N K I R L K E Q  900

2749 AATGTAGCTTTTTCACCACTTAAGAAACTGCCAAGTTTTGGACCTTATCCCCAAAAGGCCTATTCCTACCATCAAGGATGTAATAGGAAAAGCACTGCAG 2838
  901 N V A F S P L K R N C F I P K R P I P T I K D V I G K A L Q  930
                    ↑

2839 TACCTTGGAACATTTGGTGAATTGAGCAACGTAGAGCAAGTTGTGGCTATGATTGATGAAGAAATGTGTATCAACTGTGGTAAATGCTAC 2928
  931 Y L G T F G E L S N V E Q V V A M I D E E M C I N C G K C Y  960

2929 ATGACCTGTAATGATTCTGGCTACCAGGCTATACAGTTTGATCCAGAAACCCACCTGCCCACCATAACCGACACTTGTACAGGCTGTACT 3018
  961 M T C N D S G Y Q A I Q F D P E T H L P T I T D T C T G C T  990

3019 CTGTGTCTCAGTGTTTGCCCTATTGTCGACTGCATCAAAATGGTTTCCAGGACAACACCTTATGAACCAAAGAGAGGCGTACCCTTATCT 3108
  991 L C L S V C P I V D C I K M V S R T T P Y E P K R G V P L S 1020

3109 GTGAATCCGGTGTGTTAAGGTGATTTGTGAACAGTTGCTGTGAACTTCATGTCACCTACATATGCTGATCTTTTAAAATCATGATCCT 3198
1021 V N P  V  C End                                                                          1025

3199 TGTGTTCAGCTCTTCCAAATTAAAACAAATATACATTTTCTAAATAAAATATGTAATTTCAAAATACATTTGTAAGTGTAAAAAATGT 3288

3289 CTCATGTCAATGACCATTCAATTAGTGGTCATAAAATAGAATAATTCTTTTCTGAGGATAGTAGTTAAATAACTGTGTGGCAGTTAATTG 3378

3379 GATGTTCACTGCCAGTTGTCTTATGTGAAAAATTAACTTTTTTGTGGCAATTAGTGTGACAGTTCCAAATTGCCCTATGCTGTGCTCCA 3468

3469 TATTTGATTTCTAATTGTAAGTGAAATTAAGCATTTTGAAACAAGTACTCTTTAACATACAAGAAAATGTATCCAAGGAAACATTTTAT 3558

3559 CATTAAAAATTACCTTTAATTTTAATGCTGTTTCTAAGAAAATGTAGTTAGCTCCATAAAGTACAAATGAAGAAAGTCAAAAAATTATTT 3648
```

FIG. 7C

```
3649 GCTATGGCAGGATAAGAAAGCCTAAAATTGAGTTTGTAGAACTTTATTAAGTAAAATCCCCCTTCGCTGAAATTGCTTATTTTTGGTGTTG 3738

3739 GATAGAGGATAGGGAGAATATTTACTAACTAAATACCATTCACTACTCATGCGTGAGATGGGTGTACAAACTCATCCTCTTTTAATGGCA 3828

3829 TTTCTCTTTAAACTATGTTCCTAACAAAATGAGATGATAGGATAGATCCTGGTTACCACTCTTTGCTGTGCACATACGGGCTCTGACTG 3918

3919 GTTTTAATAGTCACCTTCATGATTATAGCAACTAATGTTTGAACAAAGCTCAAAGTATGCAATGCTTCATTATTCAAGAATGAAAAATAT 4008

4009 AATGTTGATAATATATTAAGTGTGCCAAATCAGTTTGACTACTCTCTGTTTTAGTGTTTATGTTTAAAAGAAATATATTTTTTGTTAT 4098

4099 TATTAGATAATATTTTTGTATTTCTCTATTTTCATAATCAGTAAATAGTGTCATATAAACTCATTTATCTCCTCTTCATGGCATCTTCAA 4188

4189 TATGAATCTATAAGTAGTAAATCAGAAAGTAACAATCTATGGCTTATTTCTATGACAAATTCAAGAGCTAGAAAAATAAAATGTTTCATT 4278

4279 ATGCACTTTTAGAAATGCATATTTGCCACAAAACCTGTATTACTGAATAATATCAAATAAAATATCATAAAGCATTTTAAAAAAAAAAAA
```

FIG. 7D

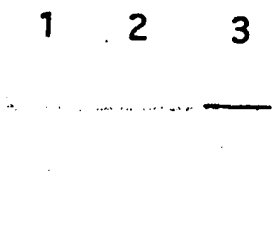

FIG. 8

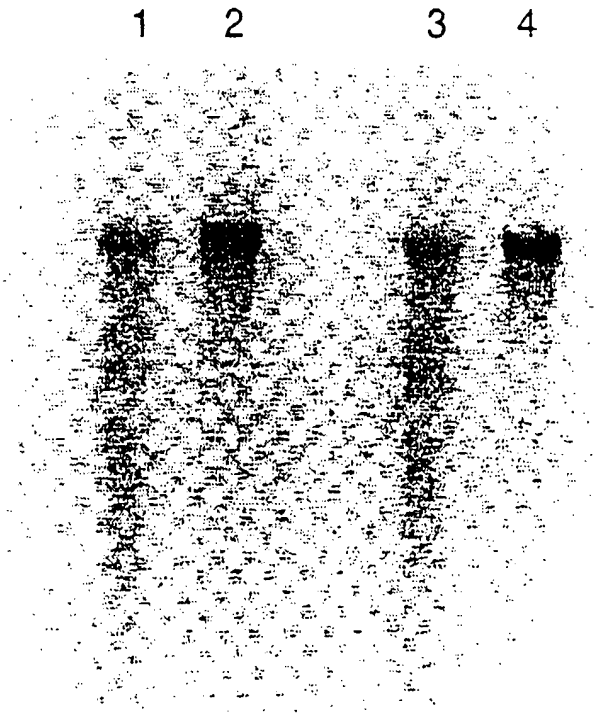

FIG. 9

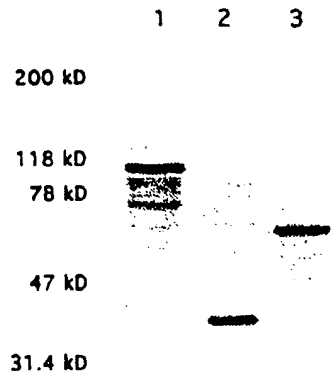

FIG. 10A

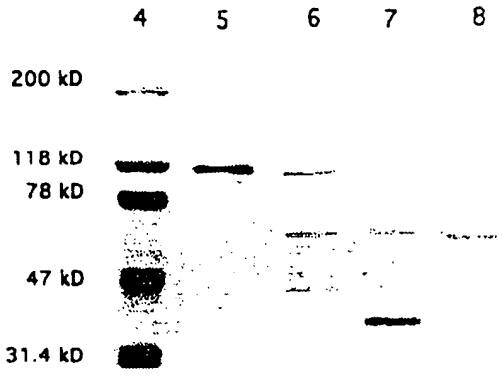

FIG. 10B

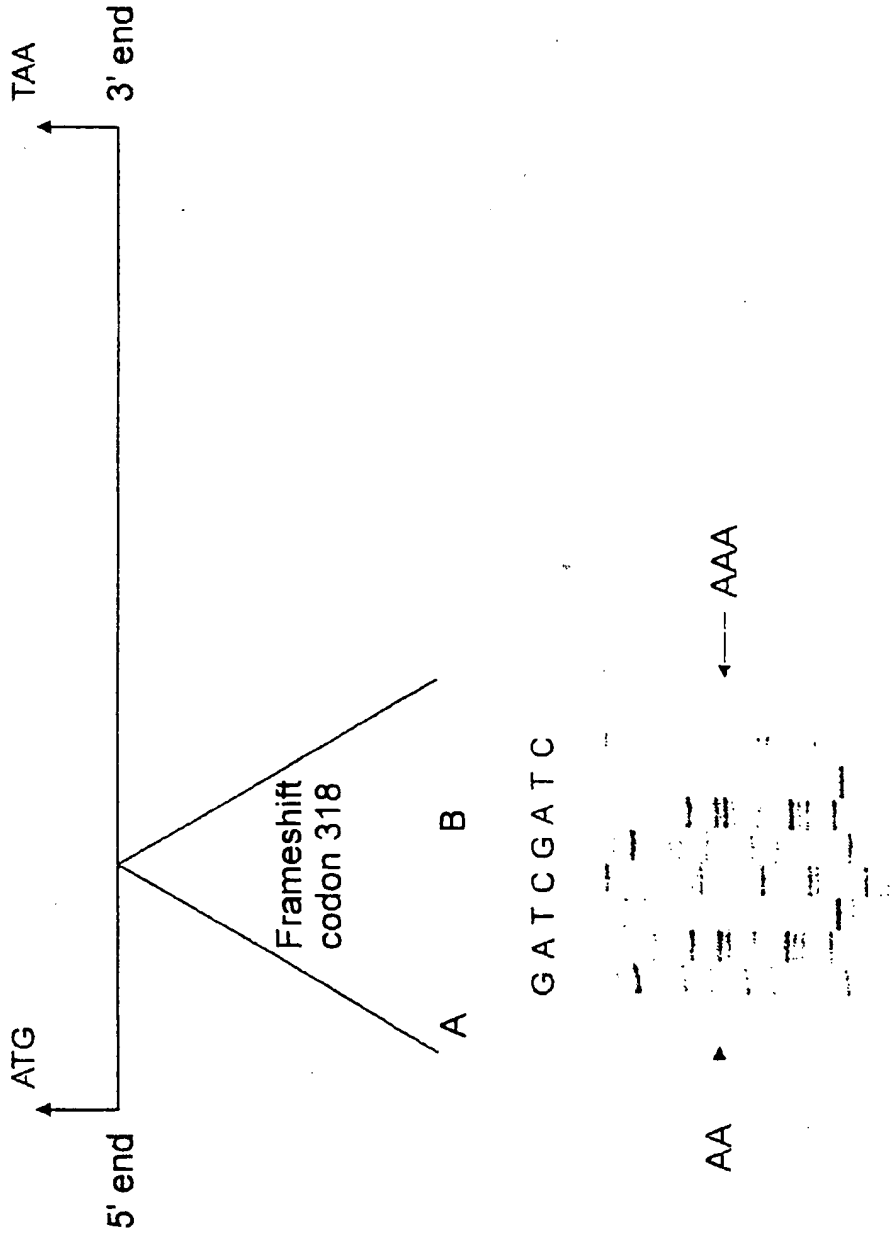

FIG.11

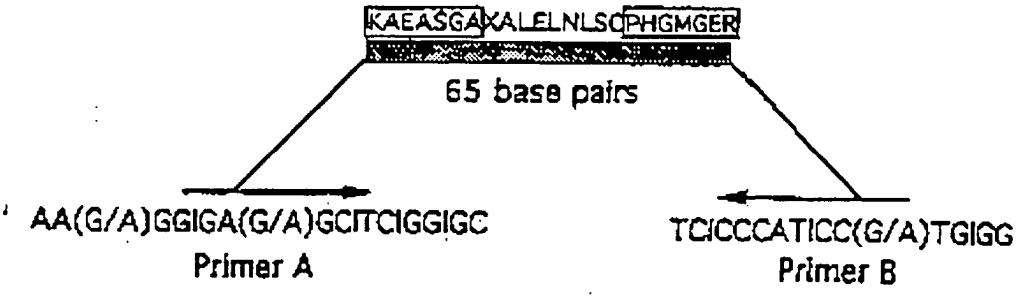
KAEASGAKALELNLSCPHGMGER
65 base pairs
AA(G/A)GGIGA(G/A)GCITCIGGIGC
Primer A
TCICCCATICC(G/A)TGIGG
Primer B

FIG. 1A

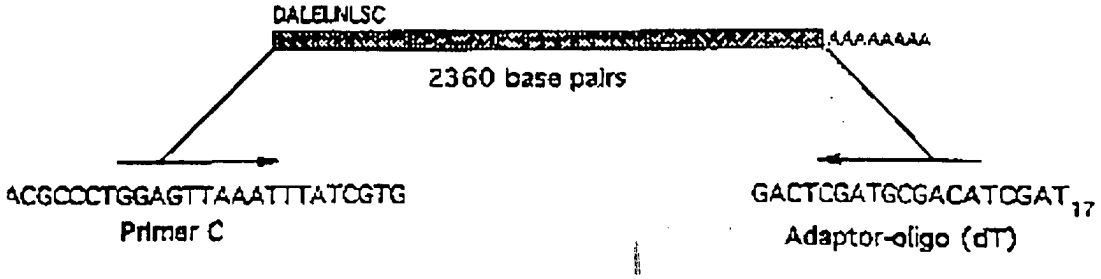
DALELNLSC
AAAAAAAA
2360 base pairs
ACGCCCTGGAGTTAAATTTATCGTG
Primer C
GACTCGATGCGACATCGAT17
Adaptor-oligo (dT)

FIG. 1B

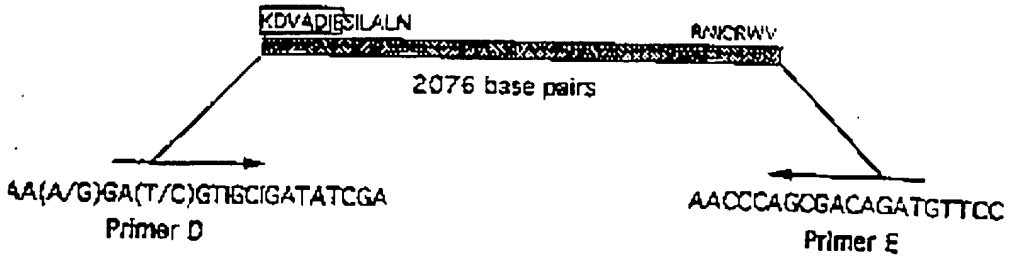
2076 base pairs
AA(A/G)GA(T/C)GTIGCIGATATCGA
Primer D
AACCCAGCGACAGATGTTCC
Primer E

FIG. 1C

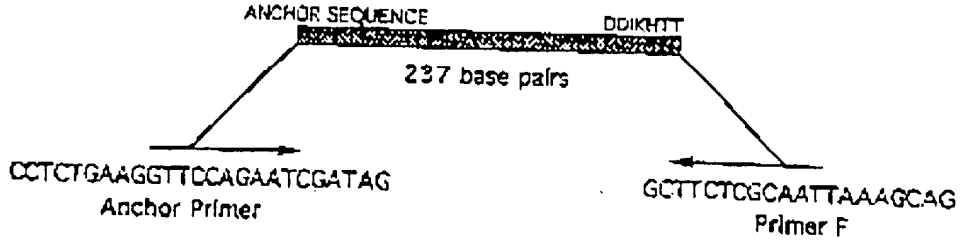
ANCHOR SEQUENCE
237 base pairs
CCTCTGAAGGTTCCAGAATCGATAG
Anchor Primer
GCTTCTCGCAATTAAAGCAG
Primer F

FIG. 1D

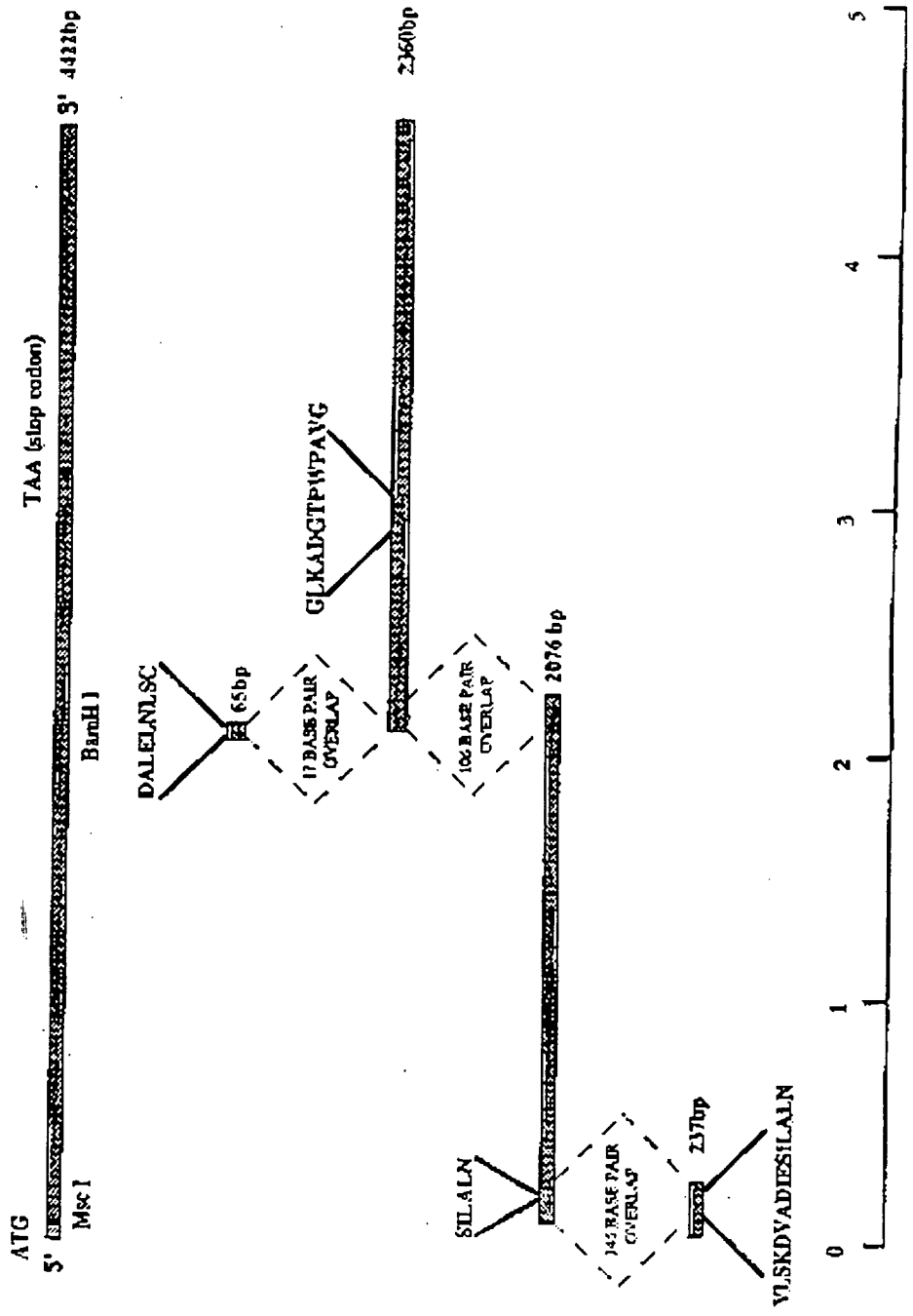
ATG
5'
Msc I
TAA (stop codon)
3' 4421bp
BamH I
DALELNLSC
65bp
17 BASE PAIR OVERLAP
GLKADGTPWPAVG
2360bp
106 BASE PAIR OVERLAP
2076 bp
SILALN
145 BASE PAIR OVERLAP
237bp
VLSKDVADTESILALN
0
1
2
3
4
5
(kb)

FIG. 2

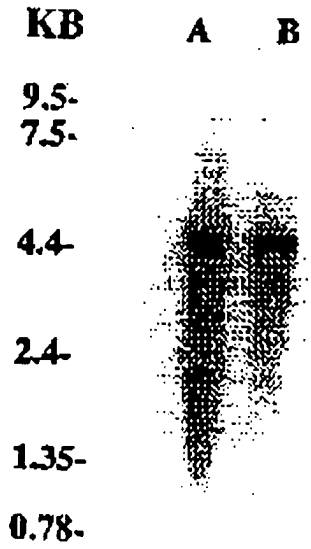

FIG. 3

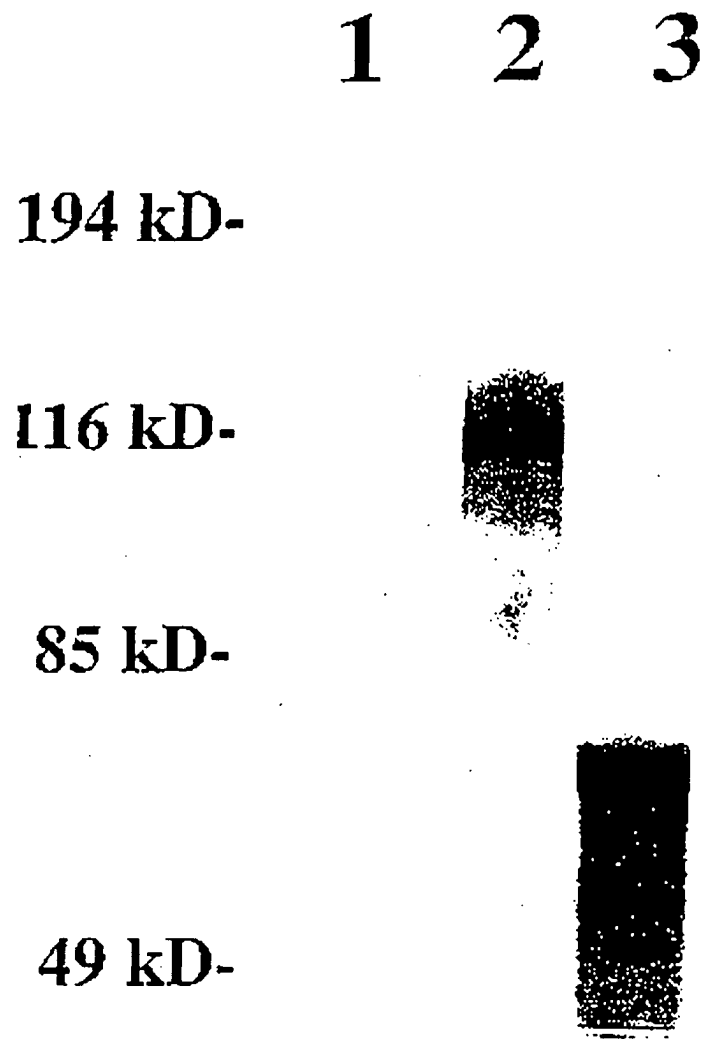

FIG. 4

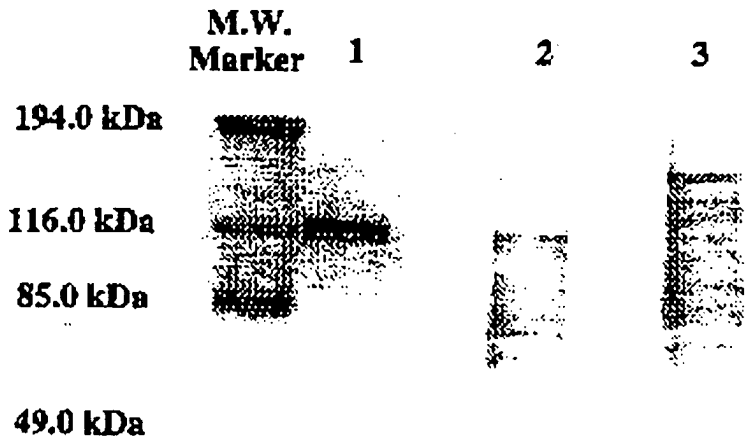

FIG. 5

FIG. 6A

FIG. 6B

FIG. 6C

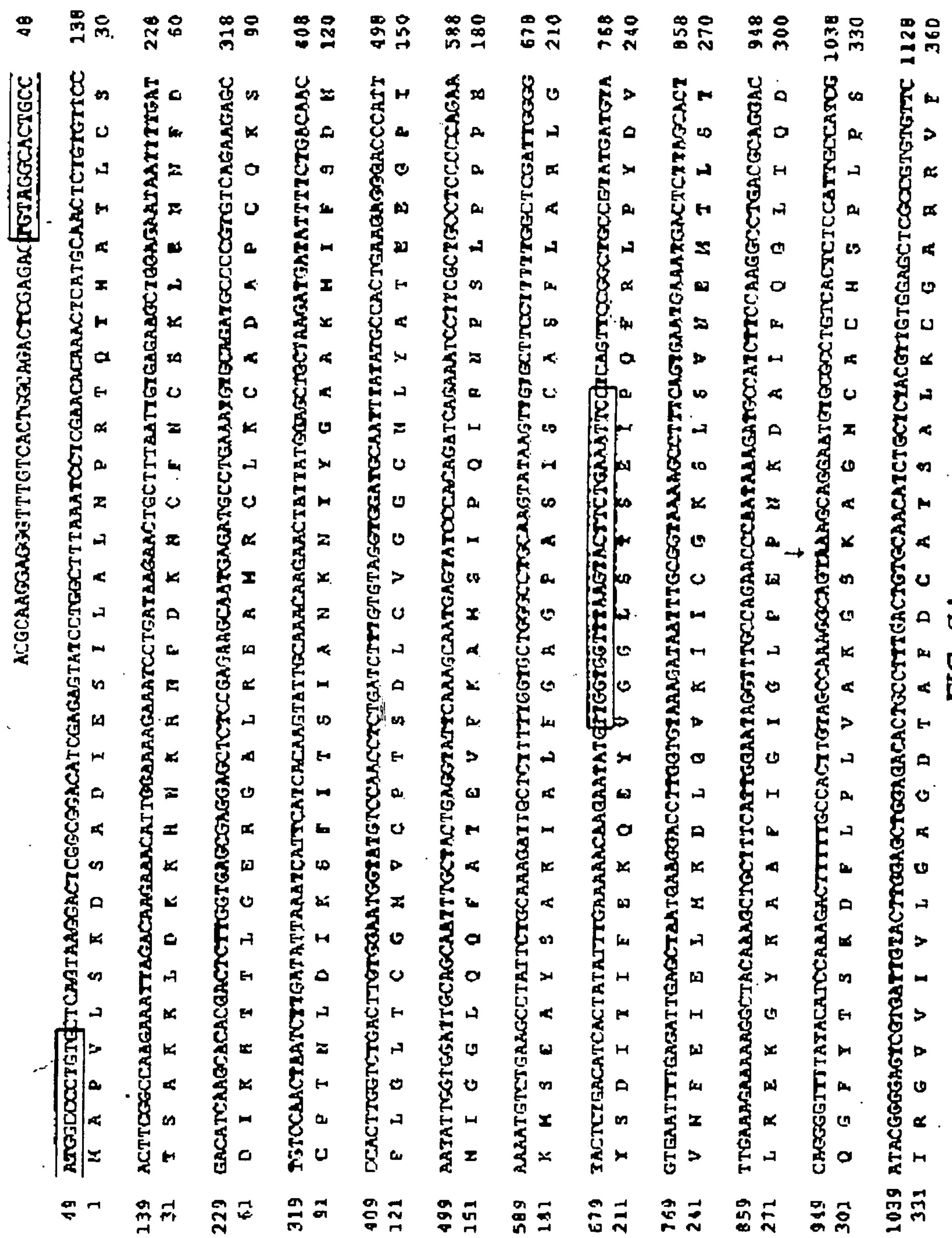

FIG. 7A

```
1129 ATCGTCTTCAGAAAAGGCTTGTTATATAAGAGCTGTCCCTGAGGAGATGGAACTTGCTAGGAAGAAAAGTGTGATTTCTGCCATTC 1218
 361  I  V  F  R  K  G  E  V  A  N  I  R  A  V  P  E  E  M  S  L  A  K  E  E  K  C  E  F  L  P  F   390

1219 CTGTCCCCACCGAAGGTTATAGTAAAAGGTGGGAGAATTGTTGCTATGCAGTTTGTTCGGACAGAGCAAGTGAACTGGAAATGGAAT 1308
 391  L  S  P  R  K  V  I  V  K  G  G  R  I  V  A  M  Q  F  V  R  T  E  Q  D  E  T  G  K  W  N   420

1309 GAAGATGAAGATCAGATGGTCCATCTGAAAGCCGATGTGGTCATCAGTGCCTTGGTTCAGTTCTGAGTGATCCTAAAGTAAAAGAGCC 1398
 421  E  D  E  D  Q  M  V  H  L  K  A  D  V  V  I  S  A  F  G  S  V  L  S  D  P  K  V  K  E  A   450

1399 TTGAGCCCTATAAATTTAACAGATGGGGTCTCCCKGAAGTAGATCCAGAAACTATGCAAACTAGTGAAGCATGGGTATTTGCKGGTGGT 1488
 451  L  S  P  I  K  F  N  R  W  G  L  P  E  V  D  P  E  T  M  Q  T  S  E  A  W  Y  F  A  G  G   480

1489 GATGTCGTTGGTTTGGCTAACACTACAGTGGAATGGATGAAAGCAGCTTCTTGTTACATTCACAAATACGTACAGTCACAA 1578
 481  D  V  V  G  L  A  N  T  T  V  E  S  V  N  D  G  K  Q  A  S  W  Y  I  R  K  I  V  Q  S  Q   510

1579 TATGGAGCTTCCGTTCTGCCAAGCCTGAACTACCCCTCTTTACATCCTATTGATCTGGTGGACATTAGTGTAGAAATGGCCGGATTG 1668
 511  Y  G  A  S  V  S  A  K  P  E  L  P  L  F  Y  T  P  I  D  L  V  D  I  S  V  E  M  A  G  L   540

1669 AAGTTTATAAATCCTTTTGGTCTGTGCTAGCGCAACTCCAGCCACCAGCACATCAATGATTCGAAGAGCTTTTGAAGCTGGATGTGGTTTT 1758
 540  K  F  I  N  P  F  G  L  A  S  A  T  P  A  T  S  T  S  H  I  R  R  A  F  E  A  G  M  G  F   570

1759 GCCCTCACCAAAACTTCTCTCTTGATAAGGACATTGTGACAAATGTTCCCCCAGAATCATCCGGGAACCACCTCTGGCCCCATGTAT 1848
 571  A  L  T  K  T  F  S  L  D  R  D  I  V  T  N  V  S  P  R  I  I  R  G  T  T  S  G  P  M  Y   600

1849 GGCCCTGGACAAAGCTCCTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCT 1938
 601  G  P  G  Q  S  S  F  L  N  I  E  L  I  S  E  K  T  A  A  Y  W  C  Q  S  V  T  E  L  K  A   630

1939 GACTTTCCAGACAACATTGTGATTGCTAGCATTATGTGCAGTTMCAATAAAATGACTGGACGGAACTTGCCAAGAAGTCTGAGGATTCT 2028
 631  D  F  P  D  N  I  V  I  A  S  I  M  C  S  Y  M  K  N  D  W  T  E  L  A  K  K  S  E  D  S   660

2029 GGAGCAGATGCCCTGGAGTAAATTATCATGTCCACATGGCATGGGAGAAAGAGGAATGGCCTGGCCTGTGGGCAGGATCCAGAGCTG 2118
 661  G  A  D  A  L  E  L  N  L  S  C  P  H  G  N  Q  E  R  G  M  G  L  A  C  G  Q  D  P  E  L   690

2119 GTGCGGAACATCTGCCGCTGGGTTAGGCAAGCTGTTCAGATTCCTTTTTTGCCAAGCTGACCCAAATGTCACTGATATTGTGMGCATC 2208
 691  V  R  N  I  C  R  W  V  R  Q  A  V  Q  I  P  F  F  A  K  L  T  P  N  V  T  D  T  V  S  I   720

2209 GCAAGAGCTGCAAGGAAGGTGGTGCCAATGGCGTTACAGCCACCAACACTGTCTCAGGTCTGATGGGATAAAATCTGATGGCACACCT 2298
 721  A  R  A  A  K  E  G  G  A  N  G  V  T  A  T  M  T  V  S  G  L  M  G  L  K  S  D  G  T  P   750
```

FIG. 7B

```
2299 IGCCAGCAGTGGGATTGCAAGCGACTACATATGGAGGAGTGTCTGGGACAGCAATCAGACCTATTGCTTTGAGAGCTGTGACCTCC 2388
 751 D P A V G I A K R R T T Y G G V S G T A I R P I A L R A V T S  780

2389 ATTGCTCGTGCTCTGCCTGGTGAATTCCCATTTGCCTACTGTGTGAAATTGGTCTTCAGTTTCTCCATAGTGGTGCT 2478
 781 I A R A L P G F P I L A T G G I D S A E S G L Q P L H S G A  810

2479 TCGGTCTCCAGTATGCAGTGCCATTCAGAATCAGGATTTCACTGTGAGTCGAGACTACAGCACTGGCCTCAAGCCCTGCTTTATCTG 2568
 811 S V L Q V C S A I Q N Q D F T V I E D Y C T G L K A L L Y L  840

2569 AAAGCATTGAAGRACTACAAGACTGGGATGGACAGAGTCCACCTACTGTGAGTCACCAGAAGGGHAACCAGTTCCACGTATAGCTGAA 2658
 841 K S I E E L Q D W D G Q S P A T Y S H Q K G K F V P R I A E  870

2659 CTCATGGACAAGAAACTGCCAAGTTTTGGACCTTATCTGGAACAGCGCAAGAAAATCATGCAGAAACAAGATTAGACTGAAAGCAACAA 2748
 871 L M D K K L P S F G P Y L E Q R K K I I A E N K I R L K E Q  900

2749 ATGTAGCTTTTTCACCACTTAAGAAACTGTTTTATCCCCAAAAGGCCTATTCCTACCATCAAGGATGTAATAGGAAAAGCACTCCAG 2838
 901 N V A F S F L K R N C F I P R R P I P T T I K D V I G K A L Q  930

2839 TACCTTGGAACATTTGGTGAATTGAGCGTGAGTATGATTGATGAGAAATGTGTATCAACTGTGGTAAATGGTAC 2928
 931 Y L G T F G E L S N V E Q V V A M I D E E M C I N C G K C Y  960

2929 ATGACCTGTAATGATTCTGGCTACCAGGCTATACAGTTTGATCAGAACCCACCTGCCCACCATACCGACACTTGTACAGGCTGTACT 3018
 961 M T C N D S G Y Q A I Q F D P E T H L P T I T D T C T G C T  990

3019 CTGTGTCAGTGTTTGCCCAATGTGCACTGCATCAAAAGGTTCCAGGACAACACCTTATGAACCAAGAGAGCGTACCCTTATCT 3108
 991 L C L S V C P I V D C I K M V S R K T P Y E P K R G V P L S  1020

3109 GTGATCCGGTGTTAAGGTGATTTGTGAACTTCATGTGCTGTGAACTTCATGTGCTACATATGCTGAACTTTTAAAATCATGATCCT 3198
1021 V N P V C End                                                                          1025

3199 TGTGTTCAGCACTTTCCAAATTAAACAAATATACATTTTCTAAATAAAATATGTAATTTCAAATACATTTGTAAGTGTAAAAAATGT 3288

3289 CTCATGTCAATCACCATTCAATTAGTGGTCATAAATAGAATAATTCTTTTCTGAGGATAGTAGTTAAATAACTGTGTGCAGTTAATTG 3378

3379 GATGTTCACTGCCAGTTGTCTTATGTGAAAATTAACTTTTTGTGGCAATTAGTGTGACAGTTCCAAATTCCCCTATGCTGTGCTCA 3468

3469 TATTTGATTTCTAATTGTAAGTGAATTAAGCATTTGAACAAAGTACTCTTTAACATACAAGAAATGTATCCAAGAAACATTTTAT 3558

3559 CATTAAAATTACCTTTAATTTTATGTTTCTAAGAAATGTAGTTGCTCCATAAAGTACAAATGAAGAAAGTCAAAAAATTATTT 3648
```

FIG. 7C

```
3649 GCTATGGCAGGNTAAGAAAGCCTAAAATTGAGTTTGTAGAACTTTATTAAGTAAAATCCCCTTCGCTGAAATTGCTTATTTTGGTGTTG 3738

3739 GATAGAGGATAGGGAGAATATTACTAACTAAATACCATTCACTACTCATGCGTGAGATGGGTGTGACAAACTCATCCTCTTTTAATGGCA 3828

3829 TTTCNCTTTAACTATGTTCCTAACAAAATGAGATAGATAGGTTACCACTCTTTGCTGTGCACATACGGGCTCTCTGCTG 3918

3919 GTTTTAATAGTCACCTTCATGATTATAGCAACTAATGTTTGAACAAAGCTCAAAGTATGCAATGCTTCATTATTCAAGAATGAAAAATAT 4008

4009 AATGTTGATAATATATTAAGTGTGCCAAATCAGTTTGACTACTCTCTGTTTTAGTGTTTATGTTTAAAAGAAATATATTTTTGTTAT 4098

4099 TATTAGATAATATTTTTGTATTTCTCTATTTTCATAATCAGTAAATAGTGTCATATAAACTCATTATCTCCTGTTCATGGCATCTTCAA 4188

4189 TATGAATCTATAAGTAGTAAATCAGAAGTAACAAATCTATTGCTATGACAAATTCAAGAGCTAGAAAAATAAAATGTTTCATT 4278

4279 ATGCACTTTTAGAAATGCATATTTGCCACAAAACCTGTAATTACTGAATAATATCAATAAATATCATAAGCATTTTAAAAAAAAAAAA
```

FIG. 7D

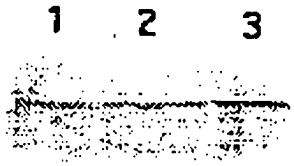

FIG. 8

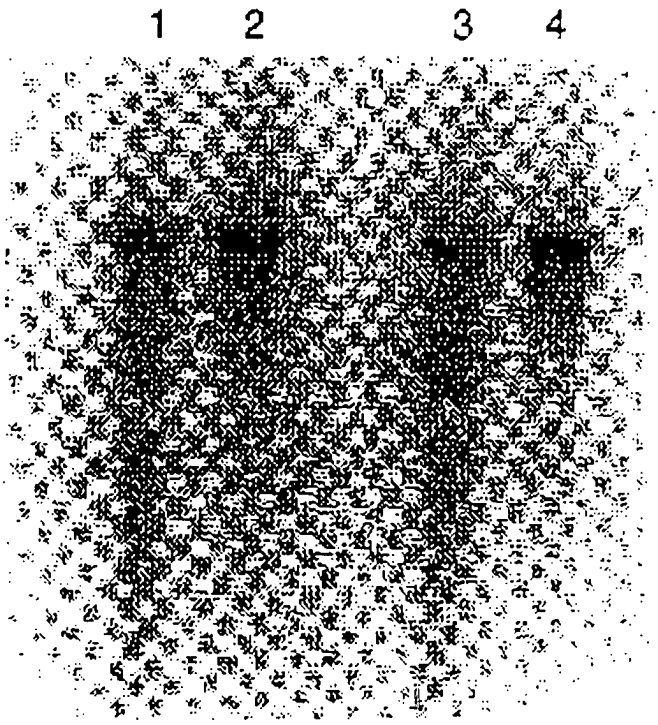

FIG. 9

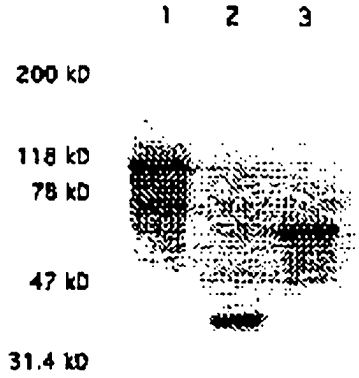

FIG. 10A

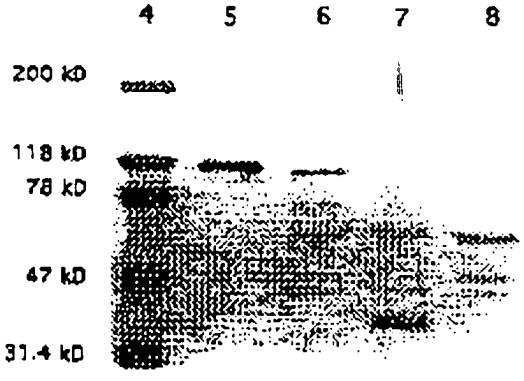

FIG. 10B

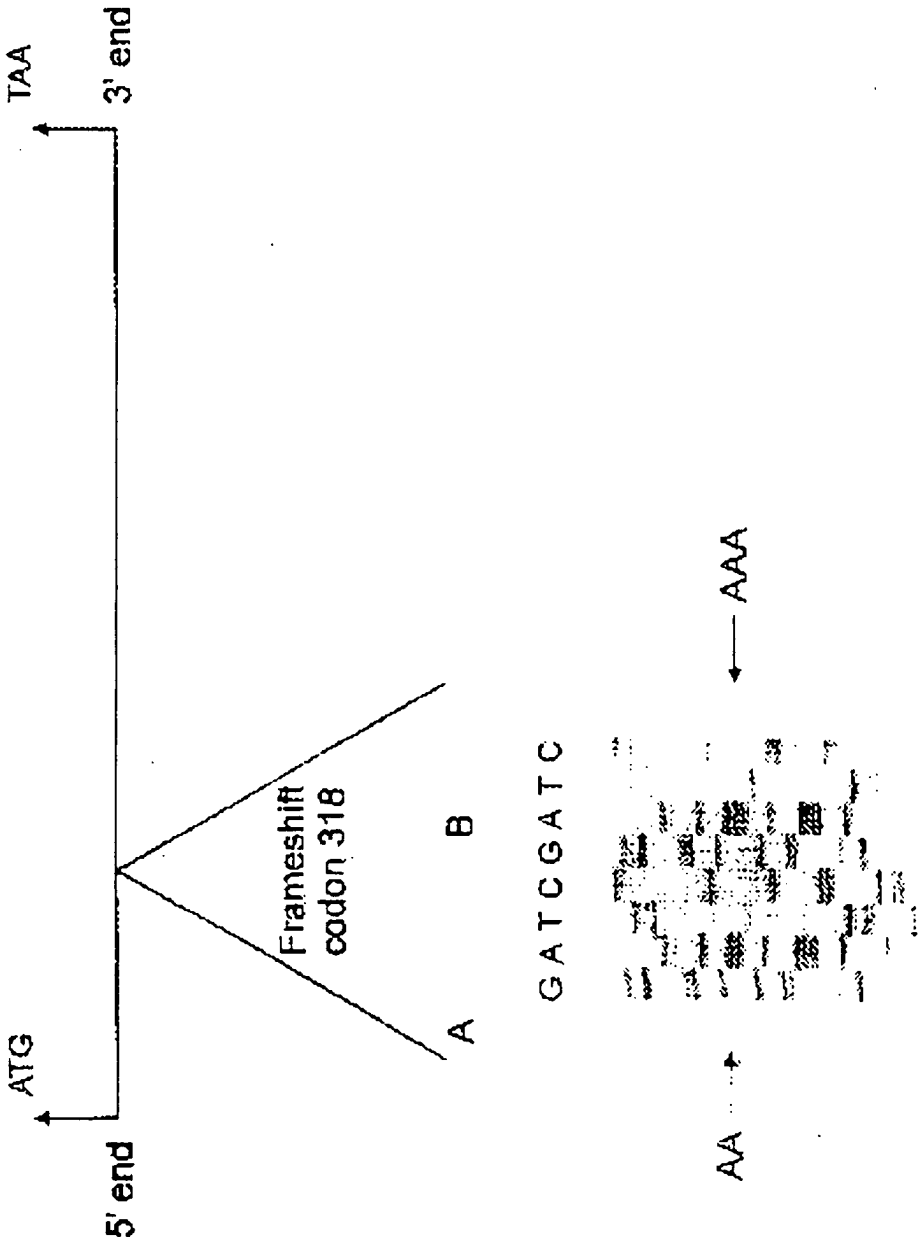

FIG.11